This guide provides an overview of techniques useful for problem in general, but it is particularly helpful for proving mathematical statements. The main sources I used to create this document are Polya's *How to Solve It* and Chartrand's *Mathematical Proofs: A Transition to Advanced Mathematics.*

Contents

xvi

HOW TO SOLVE IT

UNDERSTANDING THE PROBLEM

First.

You have to *understand* the problem.

What is the unknown? What are the data? What is the condition?

Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?

Draw a figure. Introduce suitable notation.

Separate the various parts of the condition. Can you write them down?

How to Solve It

DEVISING A PLAN

Second.

Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a *plan* of the solution.

Have you seen it before? Or have you seen the same problem in a slightly different form?

Do you know a related problem? Do you know a theorem that could be useful?

Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.

Here is a problem related to yours and solved before. Could you use it? Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?

Could you restate the problem? Could you restate it still differently? Go back to definitions.

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?

Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

How to Solve It

CARRYING OUT THE PLAN

Third.

Carry out your plan.

Carrying out your plan of the solution, *check each step*. Can you see clearly that the step is correct? Can you prove that it is correct?

LOOKING BACK

Fourth.

Examine the solution obtained.

Can you *check the result?* Can you check the argument?

Can you derive the result differently? Can you see it at a glance?

Can you use the result, or the method, for some other problem?

xvii

1 Heuristics for Problem Solving

1.1 Analogy

Definition An analogy is a comparison between two nouns.

Example To find the length of the diagonal of a rectangular prism, we can first draw an analogy to a simpler problem: finding the diagonal of a rectangle. This can in turn be analogized to a simpler problem: finding the length of the hypotenuse of a right triangle. We can see how these questions relate to each other, and by solving the simplest one, we have a model to follow; by analogy we can solve the most complicated one.

Methods

1. Using the method: When we break a problem down into simpler analogous problems, in some cases we can use only the method of solving the simpler problem to solve the more complicated problems.
2. Using the result: In some other cases, we can actually use the result of the simpler problem to solve the more complicated problem.
3. Using the method and the result: Sometimes it is possible to use both the method and the result of the simpler analogous problem.
4. Neither the method nor the result can be used: In this case, it may be worth while to reconsider the solution, to vary and to modify it till, after having tried various forms of the solution, we find eventually one that can be extended to our original problem.
5. Inference by analogy: It is desirable to foresee the result, or, at least, some features of the result, with some degree of plausibility. Thus, knowing that two OR MORE nouns (the greater in quantity, the heavier the weight of the inference) are alike in many respects, we conjecture that they are alike in one or more respect. There isn't certainty here, only plausibility. **Example:** We may know that the center of gravity of a homogeneous triangle coincides with the center of gravity of its three vertices (that is, of three material points with equal masses, placed in the vertices of the triangle). Knowing this, we may conjecture that the center of gravity of a homogeneous tetrahedron coincides with the center of gravity of its four vertices.

1.2 Auxiliary Elements

Definition An element that we introduce in the hope that it will further the solution.

Example Suppose the problem which we are trying to solve is a geometric problem, and we have recalled a related problem which we have solved before; suppose that this problem is about triangles. Yet there is no triangle in our figure. In order to make any use of the problem recollected, we must have a triangle; therefore, we have to introduce one by adding suitable auxiliary lines to our figure.

Methods

1. Having recollected a formerly solved related problem and wishing to use it for our present one, we must often ask: Should we introduce some auxiliary element in order to make its use possible?
2. Going back to definitions, we can introduce auxiliary elements by making parts of the definition explicit in our figures, equations, etc.
3. In geometry, we might introduce auxiliary lines.
4. In algebra, we might introduce auxiliary unknowns.
5. In proof-based mathematics, we might prove an auxiliary theorem whose proof we undertake in the hope of promoting the solution of our original problem.

1.3 Auxiliary Problem

Definition A problem which we consider, not for its own sake, but because we hope that its consideration may help us to solve another problem, our original problem, through familiarization with certain methods, operations, or tools which we may use afterwards for our original problem.

Example Find x, satisfying the equation $x^4 – 13x^2 + 36 = 0$. If we observe that $x^4 = (x^2)^2$ we may see the advantage of introducing $y = x^2$. We obtain now a new problem: Find y, satisfying the equation $y^2-13y+36=0$. The new problem is an auxiliary problem: we intend to use it as a means of solving our original problem. The unknown of our auxiliary problem, y, is appropriately called *auxiliary unknown.*

Methods

1. Use the method and/or the result of the auxiliary problem.
2. *Equivalent/Convertible/Bilateral Reduction:* Solving the auxiliary problem gives us the answer to the original problem (the two problems are equivalent).
3. *Chains of equivalent auxiliary problems*: in some cases, to solve problem A, we must reduce it to an equivalent problem B, which reduces to an equivalent problem C, and so on until we reach a problem whose solution is known or immediate, but that is equivalent to A. Equivalent statements can be produced using any convertible reduction (e.g. the axioms of real numbers which serve as the operational blueprint of algebra).
   (a) We must be sure that when we are passing from problem to problem, each problem is actually equivalent to every other one; if a problem becomes narrower, we lose solutions, and if it becomes wider, it loses its pertinence to the problem at hand.
4. Unilateral reduction: We have two problems, A and B, both unsolved. If we could solve A we could hence derive the full solution of B. But not conversely; if we could solve B, we would obtain, possibly, some information about A, but we would not know how to derive the full solution of A from that of B. In such a case, more is achieved by the solution of A than by the solution of B.
   (a) In some cases, we may be able to use a less ambitious auxiliary problem (B) as a stepping stone, combining the solution of the auxiliary problem with some appropriate supplementary remark to obtain the solution of the original problem (A).

1.4 Checking the Results and Argument (including intermediate results)

Methods

1. Commonsense check: Every teacher knows that students may not be disturbed when they find 16130 ft for the length of the boat and 8 years, 2 months the age of the captain who is known to be a grandfather.
2. Check the dimensions of every value derived.
3. Check by variation of the data: increase or decrease the variables in a formula solution to see if such increases/decreases affects the solution in a way that makes sense.
   (a) Convert a numerical solution into a general one with variables, and apply this check.
4. When checking the argument step by step, avoid repetition, as it is likely to become boring, uninstructive, and a strain on the attention. Where we stumbled once, there we are likely to stumble again. If we feel it is necessary to go through the whole argument again step by step, we should at least change the order of the steps or their grouping to introduce variation.
   (a) It may be better still to pick out the weakest point of the arguments and examine it first. Ask yourself did you use all the data?

1.5 Deriving the Result Differently

Definition When the solution that we have finally obtained is long and involved, we naturally suspect that there is some clearer and less roundabout solution. Try to find it. Two proofs are always better than one, as they allow you to see the object in question from multiple perspectives.

Method After having derived a formula, see if you can manipulate it so that it is expressed in a different way that has a substantially different interpretation (maybe physical) from the first formula, but still has the same overall meaning. Examine the various parts of the results, one after the other, and try various ways of considering them; we may be led finally to see the whole result in a different light, and our new conception of the result may suggest a new proof.

1.6 Using the Result or Method

Definition Having made some discovery, we should not fail to inquire whether there is something more behind it, we should not miss the possibilities opened up by the new result, we should try to use again the procedure used.

Example Given the solution to the problem of finding the diagonal of a rectangular parallelepiped, we can solve all of the following problems:

- Given the three dimensions of a rectangular parallelepiped, find the radius of the circumscribed sphere.
- The base of a pyramid is a rectangle of which the center is the foot of the altitude of the pyramid. Given the altitude of the pyramid and the sides of its base, find the lateral edges.
- Given the rectangular coordinates $(x_1, y_1, z_1), (x_2, y_2, z_2)$ of two points in space, find the distance of these points.
- Given the length, the breadth, and the diagonal of a rectangular parallelepiped, find the height.

Methods

1. Regard part of the original data from the original problem as unknown and take the unknown from the original and make it a given. Then we can use the result from the original to solve this new problem.
2. Use the results to solve more general problems via *generalization* e.g. find the diagonal of a parallelepiped, being given the three edges issued from an end-point of the diagonal, and the three angles between these three edges.
3. Use the results to solve more specific problems using *specialization*: Find the diagonal of a cube with given edge.

1.7 Carrying out a Plan

Methods

1. We may use provisional and merely plausible arguments when devising the final and rigorous argument. When the solution is sufficiently advanced, we brush aside all kinds of provisional and merely plausible arguments, and the result should be supported by rigorous argument alone.

2. We should give some consideration to the order in which we work out the details of our plan, especially if our problem is complex. We should not omit any detail, we should understand the relation of the detail before us to the whole problem, we should not lose sight of the connection of the major steps. Therefore, we should proceed in proper order.

   (a) In particular, it is not reasonable to check minor details before we have good reasons to believe that the major steps of the argument are sound. If there is a break in the main line of the argument, checking the minor details would be useless.

   (b) The order in which we work out the details of the argument may be very different from the order in which we invented them; and the order in which we write down the details in a definitive exposition may be still different.

3. The Euclidean method: In Euclid's exposition all arguments proceed in the same direction: from the data toward the unknown in "problems to find," and from the hypothesis toward the conclusion in "problems to prove." Any new element has to be correctly derived from the data or from elements correctly derived in foregoing steps. Each new element, each new assertion is examined when it is encountered first, and so it has to be examined just once; we may concentrate all our attention upon the present step, we need not look behind us, or look ahead. The very last new element whose derivation we have to check, is the unknown. The very last assertion whose proof we have to examine, is the conclusion. If each step is correct, also the last one, the whole argument is correct.

   (a) The Euclidean method is excellent to show each particular point of an argument but not so good to show the main line of the argument. The intelligent reader can easily see that each step is correct but has great difficulty in perceiving the source, the purpose, the connection of the whole argument.

4. Carrying out a plan, we check each step. Checking our step, we may rely on intuitive insight or on formal rules. Sometimes the intuition is ahead, sometimes the formal reasoning. It is an interesting and useful exercise to do it both ways; to prove formally what is seen intuitively and to see intuitively what is proved formally.

Note

1. Intuitive insight may rush far ahead of formal proof, e.g. any intelligent student can see that two straight lines parallel to the same straight line are parallel to each other. But a proof of this requires long, careful, and ingenious preparation.

2. Formal manipulation of logical rules and algebraic formulas may get far ahead of intuition, e.g. scarcely anybody is able to see that 5 planes, taken at random, divide space into 26 parts. Yet is can be rigidly proved that the right number is 26, and the proof is not long or difficult.

1.8 Condition

Definition The principal part of a "problem to find". A condition is called redundant if it contains superfluous parts. It is called contradictory if its parts are mutually opposed and inconsistent so that there is no object satisfying the condition.

Example If a condition is expressed by more linear equations than there are unknowns, it is either redundant or contradictory; if the condition is expressed by fewer equations than there are unknowns, it is insufficient to determine the unknowns; if the condition is expressed by just as many equations as there are unknowns it is usually just sufficient to determine the unknowns but may be, in exceptional cases, contradictory or insufficient.

1.9 Deriving Something Useful from the Data

Note When we have before us an unsolved problem, an open question, we have to find the connection between the data and the unknown.

Method Look at the unknown and try to think of a familiar problem having the same or a similar solution. This suggests starting the work from the unknown. Look at the data. Could you derive something useful from the data? This suggests starting the work from the data. **Note:** It appears that starting the reasoning from the unknown is usually preferable, though the alternative start, from the data, also has chances of success and must often be tried.

1.10 Restating the Problem

Definition If the problem is convoluted, try to restate it in a manner that clearly explicates the condition, the data, and the unknown.

1.11 Decomposing and Recombining

Definition You examine an object that touches your interest or challenges your curiosity, any object whose purpose and origin puzzle you, or any problem you intend to solve. You have an impression of the object as a whole but this impression, possibly, is not definite enough. A detail strikes you, and you focus your attention upon it. Then, you concentrate upon another detail; then, again, upon another. Various combinations of details may present themselves and after a while you again consider the object as a whole but you see it differently now. You decompose the whole into its parts, and you recombine the parts into a more or less different whole.

Method

1. If you go into detail you may lose yourself in details. We do not wish to waste our time with unnecessary detail; the difficulty is that we cannot say beforehand which details will turn out ultimately as necessary and which will not. Therefore, let us first understand the problem as a whole. Having understood the problem, we shall be in a better position to judge which particular points may be the most essential. Having examined one or two essential points we shall be in a better position to judge which further details might deserve closer examination. Let us go further into detail and decompose the problem gradually, but not further than we need to.
2. For “problems to find”: having understood the problem as a whole, we wish to go into further detail.

   (a) Where should we begin? In almost all cases, it is reasonable to begin with the consideration of the principal parts of the problem which are the unknown, the data, and the condition.

   (b) If we wish to examine further details, what should we do? Fairly often, it is advisable to examine each datum by itself, to separate the various parts of the condition, and to examine each part by itself.

   (c) If our problem is more difficult, we may wish to decompose it further, and to examine still more remote details. Thus, it may be necessary to go back to the definition of a certain term, to introduce new elements involved by the definition, and to examine the elements so introduced.

3. After having decomposed the problem, we try to recombine its elements in some new manner. Especially, we may try to recombine the elements of the problem into some new, more accessible problem which we could possibly use as an auxiliary problem. There are unlimited possibilities of recombination; difficult problems demand hidden, exceptional, original combinations. There are, however, certain usual and relatively simple sorts of combinations, sufficient for simpler problems, which we should know thoroughly and try first, even if we may be obliged eventually to resort to less obvious means:

   (a) Keep the unknown and change the rest (the data and the conditions).

      i. We may try to recollect a formerly solved problem of this kind: And try to think of a familiar problem having the same or a similar unknown. Failing to do this, we may try to invent one: Can you think of other data appropriate to determine the unknown? A new problem which is more closely related to the proposed problem has a better chance of being useful, therefore, keeping the unknown, we try to keep also some data and some part of the condition, and to change, as little as feasible, only one or two data and a small part of the condition. A good method is one in which we omit something without adding anything; we keep the unknown, keep only a part of the condition, drop the other part, but do not introduce any new clause or datum.

   (b) Keep the data and change the rest (the unknown and the condition).

      i. We may try to introduce some useful and more accessible new unknown from the data. The new unknown should be more easily obtained from the data than the original unknown. The new unknown should be useful, i.e. capable of rendering some definite service in the search of the original unknown. E.g., in finding the diagonal of a parallelepiped, we might introduce the diagonal of a face as a new unknown.

   (c) Change both the unknown and the data.

      i. In this case, we deviate more from our original course than in the foregoing cases. We sense the danger of losing the original problem altogether. Yet we may be compelled to such extensive change if less radical changes failed to produce something accessible and useful. Could you change the unknown or the data or both so that the new unknown and the new data are nearer to each other?

4. For "problems to prove": Having understood the problem as a whole, we should examine its principal parts. These are the hypothesis and the conclusion of the theorem that we are required to prove. We should understand these parts thoroughly; if there is need to get down to more particular points, we may separate the various parts of the hypothesis, and consider each part by itself. Then we may proceed to other details, decomposing the problem further and further. After having decomposed the problem, we may try to recombine its elements in some new manner. Especially, we may try to recombine the elements into another theorem

   (a) Keep the conclusion and change the hypothesis: Look at the conclusion and try to think of a familiar theorem having the same or a similar conclusion. If we do not succeed in recollecting such a theorem, we try to invent one: Could you think of another hypothesis from which you could easily derive the conclusion? We may change the hypothesis by omitting something without adding anything: Keep only a part of the hypothesis, drop the other part; is the conclusion still valid?

(b) Keep the hypothesis and change the conclusion: Could you derive something useful from the hypothesis?

(c) Change both the hypothesis and the conclusion: We may be more inclined to change both if we have had no success in changing just one. Could you change the hypothesis and conclusion so that the new hypothesis and the new conclusion are nearer to each other?

Submethods

1. We can introduce two or more new problems using the method above to the same affect.
2. A “problem to find” might rely on the solution to a “problem to prove”.

1.12 Definition

Definition A statement of a term’s meaning in other terms which are supposed to be well known. *Technical terms* in mathematics are of two kinds:

1. Some are accepted as primitive terms and are not defined.
2. Others are considered as derived terms and are defined in due form; that is, their meaning is stated in primitive terms and in formerly defined derived terms.

Example We do not give a formal definition of such primitive notions as point, straight line, and plane. Yet we give formal definitions of such notions as “bisector of an angle” or “circle” or “parabola”.

Method

1. When reading a problem, ensure that you have a full understanding of the definitions of all the technical terms. This understanding must also include knowledge of how to use the definition.
2. Then we can eliminate the technical terms by using their definition; by *going back to the definition,* we get rid of the technical term but introduce new elements and new relations that the definition is made of.
3. Going back to definitions is also important in checking an argument. An argument about a mathematical object must contain something about the definition of that object or some theorem about it or both. Otherwise, the conception of the object is too vague, and the argument is bad.

Submethods

1. When solving a problem, it can be useful to know not only the definition but also some theorems about the object of the problem.
2. Obtaining different definitions of the same mathematical objects can be useful, e.g. a sphere is defined as the set of points a radial distance from a single point; but it can also be defined as a circle revolving about a diameter.

1.13 Using All the Data

Method

1. In "problems to find", always ask if you have used all the data, and if you have used the whole condition.
2. In "problems to prove", always ask if you have used the whole hypothesis. This is in the case that hypothesis consists of parts, all of which are necessary to make the theorem true.
3. If we have not used all of the essential data, conditions, terms, hypotheses, etc. then our understanding of the problem at hand is incomplete.
   (a) *This is only true if the problems are "perfectly stated" and "reasonable".* Otherwise, there may be superfluous data, or maybe the condition is contradictory or redundant.
   (b) If there is doubt about whether a problem is perfectly stated and reasonable, we should ask ourselves if it is possible to satisfy the condition.

1.14 Related Problems

Definition When solving a problem, we always profit from previously solved problems, using their result, or their method, or their experience we acquired solving them. We can try to look for closely related problems that link to our present one by generalization, specialization or analogy. In "problems to prove", it is useful to ask if there we know a theorem that could be useful in solving the problem at hand.

1.15 Examining Guesses

Definition It is foolish both to accept a vivid guess as a proven truth, and also to disregard vivid guesses altogether. Guesses of a certain kind deserve to be examined and taken seriously: those which occur to us after we have attentively considered and really understood a problem in which we are genuinely interested. Many guesses show some part of the truth (even if they turn out to be wrong) and are useful in leading to a better guess.

Method Given a certain guess that we think is plausible, try to translate it into the language of mathematics and examine it critically.

1.16 Figures

Method

1. If the problem is geometric, we have to consider a figure–either in our imagination or traced on paper. It is almost always desirable to draw a figure rather than imagine it, as without paper, it is difficult to hold the pieces in our mind at the same time.
   (a) We start a detailed consideration of a geometric problem by drawing a figure containing the unknown and the data, all these elements being assembled as it is prescribed by the condition of the problem.
2. It is best to draw figures with tools in an exact manner so that false conclusions are not made. However, a good free hand drawing can suffice; there is no danger if we concentrate upon the logical connections and realize that the figure is a help, but by no means the basis of our conclusions: the logical connections constitute the real basis.

3. The order of assembling a figure doesn't matter.
4. The different parts of the figure shouldn't exhibit apparent relations not required by the problem. E.g., lines should not seem to be equal, or to be perpendicular, when they are not necessarily so.
5. In order to emphasize the different roles of different lines, you may use heavy and light lines, continuous and dotted lines, or lines in different colors. You draw a line very lightly if you are not yet quite decided to use it as an auxiliary line.
6. In nongeometric problems, drawing figures can be useful, as all sorts of problems can reduce to problems in geometry. We should try to express everything in the language of figures.

1.17 Generalization

Definition The passing from the consideration of one object to the consideration of a set containing that object; or passing from the consideration of a restricted set to that of a more comprehensive set containing the restricted one.

Method

1. In many cases, a few instances of some pattern can infer a very general pattern, e.g. the sum of cubes is a square.
2. In many cases, in trying to solve a specific problem, we may find it easier to generalize the problem and solve the generalized problem. In these cases, the invention of the general problem is actually the main achievement in solving the specific problem. Thus, in these cases, the solution of the general problem is only a minor part of the solution of the special problem.
3. When given a numerical problem, it may be advantageous to transform it into a general problem even if the problem doesn't ask; this I because it allows us to check the result in various ways, i.e. through testing common sense and dimensions.

1.18 Ask if you've seen the problem before

Definition If you've seen this problem or one similar, it can begin the *mobilization of resources,* which can be instrumental in reaching a solution.

1.19 Find a problem related to yours that has a solution

Methods

1. Let us compare this situation with the situation in which we find ourselves when we are working at an auxiliary problem. In both cases, our aim is to solve a certain problem A and we introduce and consider another problem B in the hope that we may derive some profit for the solution of the proposed problem A from the consideration of that other problem B. The difference is in our relation to B. Here, we succeeded in recollecting an old problem B of which we know the solution, but we do not know yet how to use it. There, we succeeded in inventing a new problem B; we know (or at least we suspect strongly) how to use B, but we do not know yet how to solve it. Our difficulty concerning B makes all the difference between the two situations. When this difficulty is overcome, we may use B in the same way in both cases; we may use the result or the method, or if we're lucky, we may use both the result and the method.

2. The intention of using a certain formerly solved problem influences our conception of the present problem. Trying to link up the two problems, the new and the old, we introduce the new problem elements corresponding to certain important elements of the old problem.
3. In many cases, the consideration of a formerly solved related problem leads us to the introduction of auxiliary elements, and the introduction of suitable auxiliary elements makes it possible for us to use the related problem to full advantage in solving our present problem.

1.20 Heuristic

Definition Serving to discover. The aim of heuristic is to study the methods and rules of discovery and invention.

1.21 Heuristic Reasoning

Definition Reasoning not regarded as final and strict but as provisional and plausible only, whose purpose is to discover the solution to the present problem. Heuristic reasoning is often based on induction, or on analogy.

Note We shall attain complete certainty when we shall have obtained the complete solution, but before obtaining certainty, we must often be satisfied with a more or less plausible guess. We may need the provisional before we attain the final.

1.22 Induction

Definition Induction is the process of discovering general laws by the observation and combination of particular instances.

Notes Mathematics presented with rigor is a systematic deductive science but mathematics in the making is an experimental inductive science.

In mathematics as in the physical sciences, we use observation and induction to discover general laws. But there is a difference. In the physical sciences, there is no higher authority than observation and induction but in mathematics there is such an authority: rigorous proof.

After having worked a while experimentally it may be good to change our point of view. We have discovered an interesting result but the reasoning that led to it was merely plausible, experimental, provisional, heuristic; let us try to establish it definitively using a rigorous proof.

1.23 Inventor's Paradox

Definition The more ambitious plan may have more chances of success.

Examples When passing from one problem to another, we may often observe that the new, more ambitious problem is easier to handle than the original problem. For example:

1. More questions may be easier to answer than just one question.
2. The more comprehensive theorem may be easier to prove.
3. The more general problem may be easier to prove.

1.24 Check if it is possible to satisfy the condition

We need to determine of our problem is reasonable, i.e., it has only one (not multiple) solution. Determining this question is useful if we can answer it easily. Otherwise, the trouble we have in obtaining it may outweigh the gain in interest.

For problem's to find, ask:

1. Is the condition sufficient to determine the unknown?
2. Is it insufficient?
3. Is it redundant?
4. Is it contradictory?

For problem's to prove, ask:

1. Is it likely that the proposition is true?
2. Or is it more likely that it is false?

These questions are often useful at an early stage when they do not need a final answer, but just a provisional answer or guess.

1.25 Lemma

In trying to prove a theorem A, we can suppose a theorem B is true provisionally, postponing its proof, and use it to prove A. Later, we must come back and prove B.

1.26 Look at the Unknown

Definition When solving a problem, you must not forget the reason you are working on it. For "problems to find", look at the unknown; for "problems to prove", look at the conclusion.

Methods

1. For "problems to find", it can be helpful to find a familiar problem (maybe a problem we have solved before or a problem whose solution we have) having the same unknown.
   (a) Any theorem proved in the past which is in some way related to the theorem before us has a chance to be of some service. Yet we may expect the most immediate service of theorems which have the same conclusion as the one before us.
2. For "problems to prove", it can be helpful to find a theorem having the same conclusion.
   (a) If we cannot find a formerly solved problem having the same unknown as the problem before us, we try to find one having a similar unknown. Problems of this kind are less closely related to the problem before us and, therefore, less easy to use for our purpose in general but they may be valuable guides nevertheless.

1.27 Modern Heuristic

Definition Modern heuristic endeavors to understand the process of solving problems, especially the mental operations typically useful in this process.

Note For an English algorithm containing the essence of modern heuristic, SEE Polya::How to Solve It::Dictionary::Modern Heuristic. See also PAPPUS, for an excellent description of the process of using heuristic reasoning.

1.28 Notation

Note Notation is important. The use of signs appears to be indispensable to the use of reason. The setting up of equations appears as a sort of translation, from ordinary language into the language of mathematical symbols.

Method

1. Introducing suitable notation is crucial. For each problem, choosing notation should be done carefully. The time we spend now on choosing the notation may be well repaid by the time we save later by avoiding hesitation and confusion. Moreover, choosing the notation carefully, we have to think sharply of the elements of the problem which must be denoted. Thus, choosing suitable notation may contribute essentially to understanding the problem.
2. A good notation should be unambiguous and easy to remember. It should avoid harmful second meanings, and take advantage of useful second meanings; the order and connection of signs should suggest the order and connection of things.
3. It is inadmissible that the same symbol denote two different objects in the same inquiry. However, it is not forbidden to use different symbols for the same object.
4. The sign should be easy to recognize and should immediately remind us of the object and the object of the sign. A simple device to make signs easily recognizable is to use initials as symbols.
5. Notation is particularly helpful in shaping our conception when the order and connection of the signs suggest the order and connection of the objects.
   (a) In order to denote objects which are near to each other in the conception of the problem, we use letters which are near to each other in the alphabet.
   (b) In order to denote objects belonging to the same category, we frequently choose letters belonging to the same alphabet for one category, using different alphabets for different categories. Thus, in plane geometry we use Roman capitals for points, small Roman letters for lines, and small Greek letters for angles.
   (c) In order to denote objects belonging to different categories but having some particular relation to each other, we may choose, to denote these two objects, corresponding letters of the respective alphabets, e.g. A for an angle and a for the side.
6. A notation expressing more than another may be termed more *pregnant.*
7. Just as words have second meanings, so can mathematical symbols, which aquire second meaning from the contexts in which they are often used.
   (a) e stands for the base of natural logarithms, i for the square root of -1, π for the ratio of the circumference of the circle to the diameter. It is on the whole better to use such symbols only in their traditional meaning. If we use such a symbol in some other meaning its traditional meaning could occasionally interfere and be embarrassing, even misleading.
   (b) Second meanings of the symbols can also be helpful. A notation used on former occasions may assist in recalling some useful procedure; though we must be careful to separate clearly the present meaning of the symbol from the former meaning.

1.29 Practical Problems

For advice on solving practical problems and a comparison with mathematical problems, SEE Polya::How to Solve It::Dictionary::Practical Problems

1.30 Problems to find, problems to prove

Definition The aim of a "problem to find" is to find a certain object, the unknown of a problem. The aim of a "problem to prove" is to show conclusively that a certain clearly stated assertion is true, or else to show that it is false.

Problems to find:

1. The principal parts of a "problem to find" are the unknown, the data, and the condition.

Problems to prove:

1. The principal parts are the hypothesis and the conclusion of the theorem which has to be proved or disproved.

1.31 Progress and Achievement

Method

1. In order to solve a problem, we must have some knowledge of the subject-matter and we must select and collect the relevant items of our existing but initially dormant knowledge. There is much more in our conception of the problem at the end than was in it at the beginning; what has been added? What we have succeeded in extracting from memory. In order to obtain the solution, we have to recall various essential facts, e.g. formerly solved problems, known theorems, definitions. Extracting such relevant elements from our memory may be termed *mobilization.*
2. In order to solve a problem, we must combine these facts and their combination must be well adapted to the problem at hand. This adapting and combining activity may be termed *organization.*
3. Mobilization and organization can never really be separated: working at the problem with concentration, we recall only facts which are more or less connected with our purpose, and we have nothing to connect and organize but materials we have recollected and mobilized. Mobilization and organization are two aspects of the same complex process.
4. Enriched with all the materials which we have recalled, adapted to it, and worked into it, our mode of conception changes. Our conception of the problem is much fuller at the end than it was at the outset.
5. As we make progress toward our final goal, we see more and more of it, and when we see it better we judge that we are nearer to it. As our examination of the problem advances, we foresee more and more clearly what should be done for the solution and how it should be done, e.g. that we should use a certain known theorem, or that we should consider a formerly solved problem. We do not foresee such things with certainty, only with a certain degree of plausibility. We attain complete certainty when we have obtained the complete solution.

6. What is progress toward the solution? Advancing mobilization and organization of our knowledge, evolution of our conception of the problem, increasing prevision of the steps which will constitute the final argument. We may advance steadily, by small imperceptible steps, but now and then we advance abruptly, by leaps and bounds.
7. There are some questions we should ask ourselves that are aimed directly at the mobilization of knowledge: Have you seen it before? Have you seen the same problem in slightly different form? Etc.
8. There are some questions that we should ask ourselves when we think that we have collected the right sort of material and we work for better organization of what we have mobilized: Could you use a problem related to yours and solved before? Could you use its result? Its method? Etc.
9. There are some scenarios where we think we have not collected enough material: Did you use all the data? Did you use the whole condition? Etc.

1.32 Nonmathematical Problems

All of the suggestions here apply to solving nonmathematical problems as well.

1.33 Setting Up Equations

Definition To set up equations means to express in mathematical symbols a condition that is stated in words; it is translation from ordinary language into the language of mathematical formulas.

Method

1. In some easy cases, the verbal statement splits almost automatically into successive parts, each of which can be immediately written down in mathematical symbols.
2. In more difficult cases, the condition has parts which cannot be immediately translated into mathematical symbols. If this is so, we must pay less attention to the verbal statement, and concentrate more upon the meaning. Before we start writing formulas, we may have to rearrange the condition.
3. In all cases, we have to *understand the condition, to separate the various parts* of the condition, and to *write them all down.* Then translate each piece into symbols, and combine them according to the conditions of the problem.

1.34 Signs of progress, heuristic syllogism, and plausible reasoning

Notes If we come up with a good idea, one that seems relevant, when can say that it probably will help us achieve the solution, if not being the solution itself. But we must not forget the probably, because otherwise, we stop looking for rigorous confirmation, and the guess may not be as good as we had thought. However, if you neglect heuristic conclusions altogether, you will make no progress on a problem at all.

Many times, making progress can involve brining in all pieces of the data, finding the connection between the data and the unknown, using the whole condition, imagining an analogous problem, or anything that helps us conceive of the problem in a fuller way.

There are certain mental operations typically useful in solving problems (like the ones listed above). If such a typical operation succeeds, its success is felt as a sign of progress. Thus, understanding clearly the nature of the unknown means progress. Clearly disposing the various data so that we can easily recall any one also means progress. Visualizing vividly the condition as a whole may mean an essential advance; and separating the condition into appropriate parts may be an important step forward. And so on and so forth, corresponding to each mental operation in this list.

Looking for signs of progress, however subtle, is important in determining whether one's current plan to solve a problem is fruitful or hopeless. If it is fruitful, it will be marked by an increased confidence and a faster pace. If it is not, one may become hesitant, lose courage, fail altogether, or try another plan. Signs may misguide in any single case, but they guide us right in the majority of them.

Plausible reasoning: the signs that convince the inventor that his idea is good, the indications that guide us in our everyday affairs, the circumstantial evidence of the lawyer, the inductive evidence of the scientist, statistical evidence invoked in many diverse subjects¬–all these kinds of evidence agree in two essential points:

1. They do not have the certainty of a strict demonstration.
2. They are useful in acquiring essentially new knowledge, and even indispensable to any knowledge concerned with the physical world.

Note SEE Polya::How to Solve It::Dictionary::Signs of Progress::6&7 for an explanation of heuristic syllogisms, which involves assessing the credibility of a theory based on the truthiness of what the theory implies, and an extension of heuristic syllogisms to plausible reasoning.

1.35 Specialization and Counter-example

Definition The passing from the consideration of a given set of objects to that of a smaller set, or of just one object, contained in the given set.

Method

1. In proving a theorem, if no ideas present themselves, we should try testing out special cases of the theorem, starting from the easiest case and moving towards more difficult cases to convince ourselves that the theorem is true. Doing so will likely help us understand the workings of and possibly the construction of the theorem.
2. Sometimes, in trying out special cases we will actually prove our theorem, given that it is a theorem that states that some condition is not the case. This is called *counter-example.*
3. If we test a special case, and the test shows that the object is in accordance with the general statement, we may possibly derive some hint from its examination. We may receive the impression that the statement could be true, and some suggestion in which direction we should seek the proof. Or, we may receive some suggestion in which direction we should seek the counter-example, that is, which other special cases we should test. We may modify the case we have just examined, vary it, investigate some more extended special case, look around for extreme cases, etc.
4. Extreme cases are particularly instructive, e.g. if a general statement is supposed to apply to all mammals it must apply even to such an unusual mammal as the whale. There is a good chance that we can disprove a general statement by using the extreme since these are the cases that are apt to be overlooked by the inventors of the generalizations.
   (a) If, however, we find that the general statement is verified even in the extreme case, the inductive evidence derived from this verification will be strong, just because the prospect of refutation was strong.

5. We can also use a special case auxiliary problem as a stepping stone to solve another problem. Usually, we have to introduce some other piece of information along with the auxiliary problem so that the original problem reduces to the solution of the auxiliary problem.

1.36 Subconscious Work

Description There is some limit beyond which we shouldn't continue to try to solve a problem. It is better to leave the problem until the morning, or a few days later, as many times you will have a much easier time with it due to subconscious work; you might have a bright idea almost instantly, or the pieces of the solution might come much easier.

However, we shouldn't set aside a problem we wish to come back to without having already made some progress on the problem.

You have to want the solution to a problem badly or have spent much conscious effort and tension in order for the subconscious to work on the problem.

1.37 Suggestive Contacts

Description In order to make a general claim about something, you first need to observe that there are specific cases of that generality. These are called suggestive contacts. For example, say by chance you come across the relations:

$3 + 7 = 10$, $3 + 17 = 20$, $13 + 17 = 30$

And notice some resemblance between them. It strikes you that the numbers 3, 7, 13, and 17 are odd primes. The sum of two odd primes is necessarily an even number; 10, 20, and 30 are all even. What about the other even numbers? Do they behave similarly? The first even number which is a sum of two odd primes is $6 = 3 + 3$. Looking beyond, we find that

$8 = 3 + 5$, $10 = 3 + 7 = 5 + 5$, $12 = 5 + 7$, $14 = 3 + 11 = 7 + 7$, $16 = 3 + 13 = 5 + 11$.

Will it go on forever? At any rate, the particular cases observed suggest a general statement: Any even number greater than 4 is the sum of two odd primes.

We have arrived at a conjecture. We have arrived at this conjecture by induction, indicated by particular instances.

Method

1. First, we noticed some *similarity.* We recognized that 3, 7, 13, and 17 are primes, 10, 20, and 30 even numbers, and that the three equations are *analogous* to each other.
2. Then there was a step of generalization. From the examples 3, 7. 13. And 17 we passed to all odd primes, from 10, 20, and 30 to all even numbers, and then on to a possibly general relation: even number = prime + prime.
3. We arrived so at a clearly formulated general statement, which, however, is merely a conjecture, merely *tentative.* The statement is by no means proved, it cannot have any pretension to be true, it is merely an attempt to get at the truth.
4. The conjecture has, however, some suggestive points of contact with experience, with "the facts", with "reality". It is true for the particular even numbers 10, 20, 30 and for 6, 8, 12, 14, 16.

1.38 Supporting Contacts

Description For any conjecture, you should not take it at face value; you must try to prove or disprove it. You must *test* it.

Test the conjecture with some new examples. If you find one example that contradicts the conjecture, then it is certifiably false. If you test the conjecture with an example and the conjecture holds for that example, then this can be interpreted as a *favorable sign* for the conjecture, making it *more credible,* but is not proof.

Systematically testing and verifying a conjecture (such as Goldbach's) on different cases strengthens the conjecture, renders it more credible, and adds to its plausibility. These tests for which the conjecture is verified are called *supporting contacts.*

In order to come up with a conjecture, we need to make observations that indicate, that *suggest* the conjecture. These are called *suggestive contacts.* After we invent the conjecture we begin testing and verifying it on more examples. Every test after the invention of the conjecture is called a *supporting contact.* These are two similar yet distinct concepts.

1.39 Symmetry

Definition The quality of having interchangeable parts.

Description If a problem is symmetric in some way (e.g. $yz + zx + xy$ is symmetric, since any two of the three letters x, y, z can be interchanged without changing the expression), we may derive some profit from noticing its interchangeable parts and it often pays to treat those parts which play the same role in the same fashion.

Try to treat symmetrically what is symmetrical and do not destroy wantonly any natural symmetry. However, we are sometimes compelled to treat unsymmetrically what is naturally symmetrical.

1.40 Test by Dimension

Description We test that the dimensions of the solution are correct and adhere to common sense. We may apply the test by dimension to the final result of a problem or to intermediate results, or to our own work or to the work of others, and also to formulas that we recollect and to formulas that we guess (e.g. when trying to remember the formulas for the area and volume of a sphere, we can test the dimensions of what we recollect).

The test by dimension gives no information about the values of constants, and it gives no information about the limits of validity, e.g. the scope of objects for which a formula can be applied, regardless of the dimensions.

1.41 The Intelligent Reader

Description The intelligent teacher and author should write and lecture for the intelligent reader: each step of solving a problem should be correct AND show its purpose. Each step should convey to the intelligent reader the way in which he could have come up with that step; without this explanation or clarity, the intelligent reader will think that it is not humanly possible to have invented that step.

In writing a proof, explaining how one came up with a certain step can help the reader because it can give them ideas on how they can invent similar steps in different proofs.

1.42 Variation of the Problem

Description When we have obtained the solution of a problem, our conception of the problem will be fuller and more adequate than it was at the outset. Desiring to proceed from our initial conception of the problem to a more adequate, better adapted conception, we try various standpoints and view the problem from different sides. Success depends on choosing the right aspect, on attacking the fortress from its accessible side. In order to find which aspect is the right one, we try various sides and aspects, thus varying the *problem.*

We remember things by a kind of "action by contact", called "metal association"; what we have in our mind at present tends to recall what was in contact with it at some previous occasion. Varying the problem, we bring in new points, and so we create new contacts, new possibilities of contacting elements relevant to our problem.

We cannot hope to solve any worth-while problem without intense concentration. But we are easily tired by this. In order to keep attention alive, the object on which it is directed must unceasingly change. If our work progresses, there is something to do, new points to examine, and our attention is occupied. If we fail to make progress, our attention falters, our interest fades, we get tired of the problem, and there is danger of losing the problem altogether. To escape from this danger we have to set ourselves a new question about the problem; the new question unfolds untried possibilities of contact with our previous knowledge, thus reconquering our interest by varying the problem and showing some new aspect of it.

Certain modes of varying the problem are particularly useful: SEE Definition, Decomposing and Recombining, introducing Auxiliary Elements, Generalization, Specialization, and Analogy.

1.43 Working Backwards

Description Sometimes in starting a problem, it is best to work backwards from the solution to the given; this allows us to see what steps must be taken to get from the given to the solution, but in reverse order.

SEE Polya::How to Solve It::Dictionary::Working Backwards for an example using quarts.

2 Proofs

2.1 Trivial Proofs

For a given element $x \in S$, let's recall the conditions under which $P(x) \to Q(x)$ has a particular truth value: it is true if $Q(x)$ is true or $P(x)$ is false. Accordingly, if $Q(x)$ is true for all $x \in S$ or $P(x)$ is false for all $x \in S$, then determining the truth or falseness of $P(x) \to Q(x)$ becomes considerably easier. Indeed, if it can be shown that $Q(x)$ is true for all $x \in S$ (regardless of the truth value of $P(x)$), then, $P(x) \to Q(x)$ is true. This constitutes a proof of $P(x) \to Q(x)$ is called a **trivial proof.** Accordingly, the statement

$$\text{Let } n \in \mathbb{Z}. \text{ If } x < 0, \text{ then } x^2 + 1 > 0.$$

is true and a (trivial) proof consists only of observing that 3 is an odd integer.

2.2 Vacuous Proofs

Let $P(x)$ and $Q(x)$ be open sentences over a domain S. Then $\forall x \in S, P(x) \to Q(x)$ is a true statement if it can be shown that $P(x)$ for all $x \in S$ (regardless of the truth value of $Q(x)$), according to the truth table for implication. Such a proof is called **vacuous proof** of $\forall x \in S, P(x) \to Q(x)$. Therefore

$$\text{Let } n \in \mathbb{Z}. \text{ If 3 is even, then } n^3 > 0.$$

is a true statement.

2.3 Direct Proofs

Let $P(x)$ and $Q(x)$ be open sentences over a domain S. Suppose that our goal is to show that $P(x) \to Q(x)$ is true for every $x \in S$, that is, our goal is to show that the quantified statement $\forall x \in S, P(x) \to Q(x)$ is true. If $P(x)$ is false for some $x \in S$, then $P(x) \to Q(x)$ is true for this element x. Hence, we need only be concerned with showing that $P(x) \to Q(x)$ is true for this element x. Hence, we need only be concerned with showing that $P(x) \to Q(x)$ is true for all $x \in S$ for which $P(x)$ is true. In a **direct proof** of $P(x) \to Q(x)$ for all $x \in S$, we consider an arbitrary element $x \in S$ for which $P(x)$ is true and show that $Q(x)$ as well for this element x. To summarize then, to give a direct proof of $P(x) \to Q(x)$ for all $x \in S$, we assume that $P(x)$ is true for an arbitrary element $x \in S$ and show that $Q(x)$ must be true for this element x.

2.4 Proof by Contrapositive

Suppose that we wish to prove a result (or theorem) that is expressed as

$$\text{Let } x \in S. \text{ If } P(x), \text{ then } Q(x).$$

or as

$$\text{For all } x \in S, \text{ if } P(x), \text{ then } Q(x).$$

We have seen that a proof of such a result consists of establishing the truth of the implication $P(x) \to Q(x)$ for all $x \in S$. If it can be shown that $(\sim Q(x)) \to (\sim P(x))$ is true for all $x \in S$, then $P(x) \to Q(x)$ is true for all $x \in S$. A **proof by contrapositive** of the above results is a direct proof of its contrapositive:

$$\text{Let } x \in S. \text{ If } \sim Q(x), \text{ then } \sim P(x).$$

or

$$\text{For all } x \in S, \text{if} \sim Q(x), \text{ then } \sim P(x).$$

Thus, to give a proof by contrapositive of the above results, we assume that $\sim Q(x)$ is true for an arbitrary element $x \in S$ and show that $\sim P(x)$ is true for this element x.

2.5 Proof by Cases

While attempting to give a proof of a mathematical statement concerning an element x in S, it is sometimes useful to observe that x possesses one of two or more properties. A common property that x may possess is that of belonging to a particular subset of S. If we can verify the truth of the statement for each property that x may have, then we have a proof of the statement. Such a proof is then divided into parts called **cases**, one case for each property that x may possess of for each subset to which x may belong. This method is called **proof by cases.** Indeed, it may be useful in a proof by cases to further divide a case into other cases, called **subcases.**

For example, in a proof of $\forall n \in \mathbb{Z}, R(n)$, it might be convenient to use a proof by cases whose proof is divided into the two cases

Case 1. n is even. and *Case 2.* n is odd.

2.6 Counterexample

It must be certainly come as no surprise that some quantified statements of the type $\forall x \in S, R(x)$ are false. We know that

$$\sim (\forall x \in S, R(x)) \equiv \exists x \in S, \sim R(x),$$

that is, if the statement $\forall x \in S, R(x)$ is false, then there exists some element $x \in S$ for which $R(x)$ is false, Such an element x is called a **counterexample** of the (false) statement $\forall x \in S, R(x)$. Finding a counterexample verifies that $\forall x \in S, R(x)$ is false.

We know that a quantified statement of the type

$$\forall x \in S, R(x)$$

is false if

$$\exists x \in S, \sim R(x)$$

is true, that is, if there exists some element $x \in S$ for which $R(x)$ is false. There will be many instances when $R(x)$ is an implication $P(x) \to Q(x)$. Therefore, the quantified statement

$$\forall x \in S, P(x) \to Q(x)$$

is false if

$$\exists x \in S, \sim (P(x) \to Q(x))$$

is true. The above statement can be expressed as

$$\exists x \in S, (P(x) \wedge (\sim Q(x))).$$

That is, to show that the statement $\forall x \in S, P(x) \to Q(x)$ is false, we need to exhibit a counterexample, which, in this case, is an element of $x \in S$ for which $P(x)$ is true and $Q(x)$ is false.

2.7 Proof by Contradiction

Suppose, as usual, that we would like to show that the certain mathematical statement $R : \forall x \in S, P(x) \to Q(x)$ is true. Suppose that we assume R is a false statement and, form this assumption, we are able to arrive at or deduce a statement that contradicts some assumption we made in the proof or some known fact. (The known fact might be a definition, an axiom or a theorem.) If we denote this assumption or known fact by P, then we have deduced is $\sim P$ and have thus produced the contradiction $C : P \wedge (\sim P)$. We have therefore established the truth of the implication

$$(\sim R) \to C.$$

However, because $(\sim R) \to C$ is true and C is false, it follows that $\sim R$ is false and so R is true, as desired. This technique is called **proof by contradiction.**

If R is the quantified statement $\forall x \in S, P(x) \to Q(x)$, then a proof by contradiction of this statement consists of verifying the implication

$$\sim (\forall x \in S, P(x) \to Q(x)) \to C$$

for some contradiction C. However, since

$$\sim (\forall x \in S, P(x) \to Q(x)) \equiv \exists x \in S, \sim (P(x) \to Q(x)) \equiv \exists x \in S, (P(x) \wedge (\sim Q(x))),$$

it follows that a proof by contradiction of $\forall x \in S, P(x) \to Q(x)$ would begin by assuming the existence of some element $x \in S$ such that $P(x)$ is true and $Q(x)$ is false. That is, a proof by contradiction of $\forall x \in S, P(x) \to Q(x)$ begins by assuming the existence of a counterexample of this quantified statement. Often the reader is alerted that a proof by contradiction is being used by saying (or writing) "Suppose that R is false" or "Assume, to the contrary, that R is false." Therefore, if R is the quantified statement $\forall x \in S, P(x) \to Q(x)$, then a proof by contradiction might begin with: *"Assume, to the contrary, that there exists some element $x \in S$ for which $P(x)$ is true and $Q(x)$ is false."* (or something along these lines). The remainder of the proof then consists of showing that this assumption leads to a contradiction.

2.8 Existence Proofs

In an **existence theorem** the existence of an object (or objects) possessing some specified property or properties is asserted. Typically then, an existence theorem concerning an open sentence $R(x)$ over a domain S can be expressed as a quantified statement

$$\exists x \in S, R(x) : \text{There exists } x \in S \text{ such that } R(x).$$

We know that the above statement is true provided that $R(x)$ is true for some $x \in S$. A proof of an existence theorem is called an **existence proof.** An existence proof may then consist of displaying or constructing an example of such an object or perhaps, with the aid of known results, verifying that such objects must exist without ever producing a single example of the desired type.

2.9 Disproving Existence statements

Let $R(x)$ be an open sentence where the domain of x is S. We have already seen that to disprove a quantified statement of the type $\forall x \in S, R(x)$, if suffices to produce a counterexample (that is, an element x in S for which $R(x)$ is false). However, disproving a quantified statement of the type $\exists x \in S, R(x)$ requires a totally different approach. Since

$$\sim (\exists x \in S, R(x)) \equiv \forall x \in S, \sim R(x),$$

it follows that the statement $\exists x \in S, R(x)$ is false if $R(x)$ is false for *every* $x \in S$. Let's look at some examples of disproving existence statements.

2.10 Uniqueness

An element belonging to some prescribed set A and possessing a certain property P is **unique** if it is the only element of A having property P. Typically, to prove that only on element of A has property P, we proceed in one of two ways:

1. We assume that a and b are elements of A possessing property P and show that $a = b$.

2. We assume that a and b are distinct elements of A possessing property P and show that $a = b$.

2.11 Induction

For a fixed integer m, let $S = \{i \in \mathbb{Z} : i \geq m\}$. For each $n \in S$, let $P(n)$ be a statement. If

1. $P(m)$ is true and

2. $\forall k \in S, P(k) \to P(k+1)$ is true,

then $\forall n \in S, P(n)$ is true.

To perform a proof by induction, follow these steps:

1. **State that the proof uses induction.** This immediately conveys the overall structure of the proof, which helps the reader understand your argument.
2. **Define an appropriate predicate** $P(n)$. The eventual conclusion of the induction argument will be that $P(n)$ is true of all nonnegative n. Thus, you should define the predicate $P(n)$ so that your theorem is equivalent to (or follows from) this conclusion. Often the predicate can be lifted straight from the proposition that you are trying to prove. The predicate $P(n)$ is called the *induction hypothesis.* Sometimes the induction hypothesis will involve several variables, in which case you should indicate which variable serves as n.
3. **Prove that** $P(0)$ **is true.** This part of the proof is called the *base case* or *base step.*
4. **Prove that** $P(n)$ **implies** $P(n+1)$ **for every nonnegative integer** n. This is called the *inductive step.* The basic plan is always the same: assume that $P(n)$ is true and then use this assumption to prove that $P(n+1)$ is true. These two statements should be fairly similar, but bridging the gap may require some ingenuity. Whatever argument you give must be valid for every nonnegative integer n, since the goal is to prove the implications $P(0) \to P(1), P(1) \to P(2), P(2) \to P(3)$, etc. all at once.
5. **Invoke induction.** Given these facts, the induction principle allows you to conclude that $P(n)$ is true for all nonnegative n. This is the logical capstone to the whole argument, but it is so standard that it's usual not to mention it explicitly.

2.12 Principle of Strong Induction

(**The Strong Principle of Mathematical Induction**) For a fixed integer m, let $S = \{i \in \mathbb{Z} : i \geq m\}$. For each $n \in S$, let $P(n)$ be a statement. If

1. $P(m)$ is true and
2. the implication

   $$\text{If } P(i) \text{ for every integer } i \text{ with } m \leq i \leq k, \text{ then } P(k+1).$$

   is true for every integer $k \in S$.

2.13 Well Ordering Principle

The *Well Ordering Principle* is the following: Every nonempty set of nonnegative integers has as smallest term.

1. Define the set, C, of *counterexamples* to P being true. Namely, define

   $$C ::= \{n \in \mathbb{N} | P(n) \text{ is false}\}.$$

2. Use a proof by contradiction and assume that C is nonempty.
3. By the Well Ordering Principle, there will be a smallest element, n, in C.

4. Reach a contradiction (somehow)-often by showing how to use n to find another member of C that is smaller than n. (This is the open-ended part of the proof task.)
5. Conclude that C must be empty, that is, no counterexamples exist. QED

2.14 Invariant

1. Define $P(t)$ to be the predicate that some property holds immediately after step t.
2. Show that $P(0)$ is true.
3. Show that
$$\forall t \in \mathbb{N}. P(t) \rightarrow P(t+1).$$
namely, that for any $t \geq 0$, if a property holds immediately after step t, it must also hold after the following step.