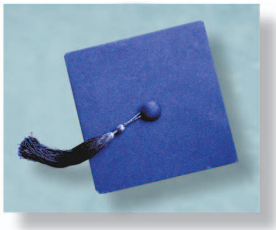


Educator's Corner

Learning Methods of Problem Solving

■ **Madhu S. Gupta**

©MASTER SERIES

There is an old well-known story about a novice truck driver driving a tall truck who misjudged the clearance below an overhead railway bridge. In trying to pass through, the truck got stuck under the bridge and thereafter would not drive in either forward or reverse gear. With the traffic blocked, a crowd began to gather around, and many people tried to help by pushing the truck but could not dislodge it. Suggestions poured in from the crowd, such as calling in a bulldozer, an army tank, or a herd of elephants to push the truck out. As time went by, calls for help were made to the traffic police, the city transportation engineers, and the railway engineers, but to no avail. The crowd began to offer more drastic ideas, ranging from breaking down the truck and digging up the road to dismantling the bridge. People laughed when a schoolboy came forward and said he might be able to help, and they dared him to try. So he walked up to the truck, and let the air out of the tires, which lowered it enough to drive out.

One lesson deducible from the story is that problems can usually be solved in multiple ways, some more efficient than others. In order to come up with an efficient method for the solution of a problem, one must first be aware of the alternative ways for solving a problem. The present article is concerned with outlining and illustrating some commonly used methods of problem solving, and with ways of helping students acquire awareness of such methods. In particular, it suggests incorporating in the classroom work multiple methods of problem solving, and a deliberate and explicit discussion of the problem solving methods employed for problem solving, as opposed to focusing solely on the solutions to particular problems. Among other potential benefits, such an approach could lead to a greater awareness of the process of problem solving among the learners.

Such learning is particularly pertinent to engineering, which is a profession of problem solving, and to engineering education, which relies heavily on problem solving as a vehicle for learning, and for numerous other purposes including learning assessment and content management, as described in [1]. The importance of problem solving skills is universally recognized, but teaching them remains a challenge in engineering education. For as long as anyone can remember, new instructors have always been surprised by the students' inability to solve simple problems, while experienced instructors concerned about the students' problem solving skills have

Madhu S. Gupta is with San Diego State University, San Diego, CA 92182, USA. mgupta@mail.sdsu.edu

Digital Object Identifier 10.1109/MMM.2008.927643

long searched for ways to improve those skills.

The significance of this knowledge can be put in proper perspective in terms of the revised Bloom's taxonomy of cognitive abilities described in these pages [1]. The taxonomy classifies a learner's knowledge of a subject into four types: the factual, conceptual, and procedural knowledge of the discipline, and the knowledge about problem solving strategies, called metacognitive knowledge. The teaching of problem solving methods is one element of education that is aimed at the fourth knowledge category of metacognitive knowledge.

How Are Problems Solved?

To discuss how an instructor can help the learners in learning problem solving, we begin by examining how a learner might go about solving a problem and what barriers may be encountered in the process. Although there is no single systematic procedure or template for solving every problem, the process of solving a problem can nevertheless be approximately outlined, or at least caricatured, by the following generic three-stage progression.

a) Interpretation

This stage concerns the elucidation and understanding of the situation described in the problem, which often requires the learner to recast the problem in a form that corresponds with that learner's own internal knowledge framework. One of the principal activities carried out in this stage is problem abstraction, which involves selective extraction (by ignoring extraneous details), synopsizing (by retaining only the essential or minimum information), as well as formalizing the representation (e.g., description at a theoretical, conceptual or analytical level), so as to help focus the attention and overcome the processing limitation imposed by the limited capacity of human short-term memory. One result of this stage is to establish the scope of problem by identifying its starting point and the final destination. A second outcome of this stage is the extraction of useful data from the given information, e.g., by suitable interpretation of verbal, numerical, tabular, or graphical information, which serves as the fuel that propels the problem solving process from the starting point to the finish line.

b) Procedurization

This stage is concerned with developing a method of solution applicable to, and suitable for, the problem at hand. The learner can either start from the given problem and reformulate it in a form to which a familiar method can be applied, start with a known method of problem solving and try to adapt it to the problem at hand, or possibly transform both the method and the problem alternately so as to match them. In the process, the learner may reformulate or re-express the problem in alternative ways, decompose the problem into a composite of multiple canonical problems, adopt a viewpoint that permits the problem to be construed as a variant of another known problem, or employ any one of the numerous techniques of problem solving. Because the process involves transforming the problem to a recognizable form, recall of the known methods of problem solution, and adaptation of a method to the problem at hand, the learner's repertoire of problem solving methods plays a significant role. The problem solver may arrive at the method of solution by a variety of ways, such as recall, search, selection, or synthesis of a procedure based on experience, trial, or intuition. The primary outcome of this stage is a problem solving procedure, i.e., the delineation of a path from the given information to the desired result.

c) Implementation

This stage consists of carrying out the selected method, procedure, or routine to solve the problem and then assessing its outcome to decide if further effort (e.g., diagnosis, error detection, or selection of an alternative method of solution) is needed.

As described above, the three steps appear to be sequential, but that holds only at the gross level. Work at any stage can require repeated returns to an earlier stage for a variety of reasons; for example, difficulty or incongruence encountered at a later stage may require a return to revise how some information was construed or modify the approximation employed. Alternatively, if at some stage, the problem is resolved into several subproblems, each one of them may require another pass through the same three stages for its solution.

Barriers to Problem Solving

When a learner is unable to solve a problem, the source of the learner's difficulty can be identified through various techniques, such as requiring the learner to think aloud or interrogating him with probing questions that can resolve and isolate the cause. In the first and third (interpretation and implementation) stages of problem solving, the source of difficulties is usually traceable to the learner's discipline-specific knowledge (of the factual, conceptual, and procedural variety), or the absence thereof. For example, an inability to interpret the problem statement, its constraints, or the alluded situation, may result from a lack of basic factual knowledge (e.g., terminology or conventions customary in the discipline); unfamiliarity with relevant physical principles (such as the consequences and use of the laws of conservation and causality); or an inability to carry out some discipline-specific procedure (like deduction of some parameters from others). Similarly, the difficulties at the implementation stage might arise from the student's unfamiliarity with a mathematical result (such as Schwartz inequality) or procedural skill (e.g., matrix manipulations); or an inability to exploit known information (such as symmetry, or linear superposition) for simplifying the problem. Once the source of difficulties has been identified, the remedial action needed is usually apparent.

By contrast, the hurdles encountered at the procedurization stage can have a very different source. Once a problem has already been interpreted and understood in a learner's framework, developing a procedure for its solution can be difficult even if all the requisite elements of factual, conceptual, and procedural knowledge are available. Novices often appear to be clueless about how to proceed, and are hindered by such difficulties as the absence of a basis for selecting a problem solving approach from among multiple contenders; the inability to plan out the task of problem solving; inadequate resourcefulness, adaptiveness, or persistence in the face of early failures in a trial-and-error process; or simply a lack of comfort, courage, or confidence in making reasonable approximations, assumptions, and idealizations that enable extraction of useful data from the problem statement. It is noteworthy that the hurdles at the procedurization stage often transcend discipline-specific knowledge, and relate to such issues as the use of problem solving strategy and the learner's personal traits.

Why Are Problem Solving Methods Usually Ignored in the Classroom

Professionals and subject-matter experts have the ability to draw upon a sizable repertoire of problem solving methods, and moreover, have a highly developed sense of where to deploy those methods. But when an instructor solves problems for students, the focus is often on the subject matter content or answers, and not on the method. As a result, the expert knowledge of what methods are available, where they are useful, and the manner in which the instructor decided how to proceed, is usually left out of the discussion. Because the problem solving process is rarely discussed, for many learners it remains shrouded in mystery.

Most problem solving methods, like difference reduction, means-ends analysis, and the use of analogies, are generic and not specific to a discipline; thus the methods used in engineering course work are no different from those encountered by learners in their earlier educational experiences such as beginning mathematics and physics courses. Moreover, the logic behind the methods is fairly elementary, and likely well-known to the students. As a result, many instructors may feel that instruction in problem solving methods per se is not within the scope of their responsibility or courses. This perception is further reinforced by the fact that the instructors themselves carry out expert-level problem solving in their own disciplines without explicitly thinking about the generic problem solving methods and strategies being employed. This may be the reason why few instructors carry out an explicit instruction in problem solving methods. Textbooks, worked example problems, instructors' solutions manuals accompanying the textbooks, course Web sites, class notes and handouts, and other instructional materials show little evidence that problem solving methods are being explicitly taught.

Is There a Need to Develop an Awareness of Problem-Solving Methods?

Some learners can, and do, learn to recognize problem solving methods on their own without a deliberate intervention on the part of the instructor, but many remain poor problem solvers as evidenced by a variety of forms of evidence: self-report during problem solving, performance on rephrased or marginally novel problems following an earlier exposure to similar problems, and perceived difficulty of assigned problems.

One of the protocols used by learning theorists in exploring human cognition and problem solving is that of self-report, in which a learner is asked to think aloud while attempting to solve a problem. Results show that most students are clueless about the problem solving methods; when asked what they are doing, they typically report something vague, like "I am solving a problem," because they are not consciously aware of a systematic approach to problem solving and tend to rely heavily on trial-and-error. In the absence of an awareness of the underlying systematic method, each problem seems to be unique that appears to require a distinct approach; this imparts an amorphous character to the task of learning, gives it the appearance of gathering a bag of tricks, and makes it overwhelming.

The students' inability to deploy previously seen methods is commonplace, the subject of much study in learning theory, and the evidence that past exposure is not sufficient for success in problem solving. All instructors will agree that the purpose of instruction is for students to learn how to solve problems other than those solved before; the process whereby this occurs is called transfer of learning [2]. Decades of work on transfer of learning has shown that transfer is limited to narrow domains, if not subject matter specific. Thus it is not enough for the student to have come across a method of solution in an earlier problem or course; the likelihood of that method being invoked is greatly increased if the student had deployed that method in the specific context of the discipline or problem type at hand. While parts of their knowledge of the problem solving method may transcend the disciplinary boundaries, other parts relate closely to the discipline, so that there is value in assisting the learners to develop a conscious awareness of the problem solving methods pertinent to a given subject, in the context of that subject matter.

One side benefit of exposing the students to problem solving strategies is to dispel some of the false and fatalistic notions carried by many students, e.g., that each problem can be solved only by a unique approach, which depends on the type of problem and must be known in advance; that given a problem, either they immediately know how to solve it or they don't, so there is no point in making efforts to look for a solution; or that the solution of an unfamiliar problem requires a divine inspiration, rather than a systematic search.

Teaching Problem Solving Methods

Perhaps the most prominent distinction between a novice problem solver and one who is an expert in the discipline, is in their problem solving strategies [3], [4]. While experts deploy efficient strategies almost effortlessly, and may not even be consciously aware of having deployed them, a novice constantly struggles with the question "what do I do next?" and make inefficient choices. Before a strategic knowledge about problem solving can be developed, the learners must first become consciously aware of the various problem solving methods that might be available. The instructor can help the learners gain this awareness in several ways.

First, an instructor can select the problems to be solved (particularly those to be solved by the instructor) to illustrate not only the subject matter but also the method of solution. This means selecting and framing the illustrative problems not only to address the cognitive abilities from various levels of Bloom's hierarchy of abilities [1], but also to illustrate the many different methods of problem solution.

Second, for each such illustrative problem exemplifying a problem solving method, it is helpful to

- identify and clarify what method is being employed when solving a problem
- give it a name, so that it can be brought to the conscious level, identified, referred, discussed, and retrieved from the memory more easily
- describe the steps in the method, and demonstrate it through its actual use in problem solving
- discuss the characteristics, domain of applicability, strengths, and limitations of each method.

Third, to convey the value of learning a variety of problem solving methods, it is helpful to demonstrate that, for a given problem, some problem solving methods can confer an advantage over others; thus a method may be more efficient than others in the required level of time, effort, and skill, or be superior in some other respect such as the resulting accuracy, form of answer, and possibility of error. At times, a method might even be preferable for reasons beyond the solution of the problem at hand, for example if it affords a particularly helpful insight, or has a broader applicability or generalizability.

Fourth, in addition to exposure and familiarity, the learn-

TABLE 1. Ten problem solving methods commonly used with elementary engineering problems.

1) Organizing Known Information	• Systematization in ways that are suggestive, revealing, or at least resolve any ambiguity • Graphical or geometrical representation of data • Hierarchical or tabular arrangement of information • Classification, clustering, or sorting of data based on different features
2) Use of a Standardized Routine	• A handbook result, rule-of-thumb or ready-made formula • Known algorithm or canned software • Tables, charts, nomograms or other graphical aids
3) Forward Logical Deduction	• Step-by-step planning towards desired goal, starting from known information, definitions, and fundamental principles • Focus on consequences deducible from information available at each intermediate step
4) Backward Reasoning	• Working backwards, starting from the desired result or end, and going towards known information • Successive reduction of goals into prerequisite goals
5) Search for Pattern or Trend	• Identifying commonalities, symmetries, or correlations • Identification of underlying canonical, elementary, or building-block problems • Abstracting trends and sensitivities resulting from parametric variation
6) Partitioning and Synthesis	• Subdivision of the problem into subproblems that have known solutions, and that can be stitched together • Subdivision of the system, the problem, or the goal, to reduce the scope or complexity of the problem, or until it is amenable to solutions by available means
7) Simplification	• Deduction of a special, simpler, limiting, asymptotic, idealized, or extreme case of the problem to solve • Extrapolation, generalization, or intelligent guessing based on the solution of a simpler problem
8) Exhaustive Enumeration	• Listing, and then evaluation, of all choices or combinations in the universe of possibilities • Creating a tree of alternatives, followed by sorting and elimination of alternatives based on some criteria.
9) Change of Goals or Problem	• Change of variables, assumptions, or model • Adopting a different viewpoint • Using analogies or metaphors to define an analogous, parallel, or equivalent problem • Transferring old solutions onto new problems
10) Simplification by Approximation	• Change of problem specifications to make it tractable • Ignoring minor details (as in order-of-magnitude estimation) • Correction of, or extrapolation from, approximate solution; successive approximation, iterative refinement

ers will need confidence, persistence, and willingness to be adaptive in problem solving. This can be enhanced through prior experience with actual use of the methods. Consequently, the instructor will need to provide opportunities for the learners to employ the various methods of problem solving, through examples, assignments, and drill, so as to gain sufficient facility in them, and develop the confidence to utilize them.

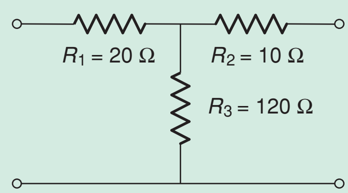
The Problem-Solving Methods

A representative list of problem solving methods that are commonly used in engineering is shown in Table 1. Although listed here in the context of engineering problem solving, the methods are generally applicable to a broad spectrum of problems in numerous disciplines and domains [5]–[7]. As a result, many of them would not be novel for the learners; but presenting a compilation of the methods that the students may already be using without explicitly recognizing and labeling them as such helps bring the methods to the conscious level. The principal purpose of exposing the students to multiple problem solving methods is to help seed the thinking required in the procedurization step of solving a new problem. This helps create not only a larger arsenal of problem solving tools to draw from, but also a richer set of links among the items of knowledge organized in the learner's memory.

A problem solving method is best discussed through its use on a sample problem that can illustrate its domain of applicability or suitability, strengths and weaknesses, and the type of information it requires and produces. Although each method may excel with a different sample problem, and a more challenging sample problem can provide a more convincing demonstration of the power of some of the solution methods, the methods will be demonstrated here by applying them to a single elementary problem. Using a simple problem, that is already expressed in an idealized and symbolic form and does not require any modeling prowess, has the advantage that the solution method, which is the object of the present discussion, is not obscured by the extraneous details found in complex problems. Using a single problem to illustrate different methods of solution further emphasizes that a given problem can be solved in a variety of ways, and that the applicability of a method is not limited to a special type of problems; indeed, a combination of multiple methods might often be the most successful approach.

It is rare that a problem can be solved by every method listed in Table 1, but a number of the methods can be exemplified by the following elementary problem from the field of microwave network theory.

Given a linear two-port network consisting of a lumped asymmetric-T resistive network shown in the following figure, the problem is to determine its scattering matrix, defined with respect to a reference impedance of 50 Ω. Reciprocity of the network ensures $S_{21} = S_{12}$.



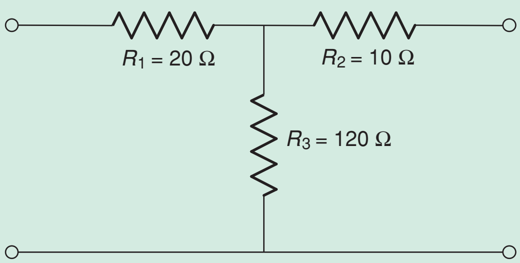
The selected problem is a typical classroom problem, which might appear to be contrived because it has a very low complexity, thus making it exactly solvable with paper and pencil. In industrial practice (or real life, as the practitioners like to call it), a problem such as this one would be solved by computer-aided design (CAD) tools, using a commercially-available software package that allows the entry of the network topology and element values via a nodal description, and computes the desired response functions. While that would indeed be the quickest method with the greatest flexibility (e.g., by permitting alterations in network, or expressing the results in different forms) and the lowest of risk of error, its purpose would usually be to get the answer, not learn a method of problem solving. What is learned by solving any given problem, including the above asymmetric resistive-T network problem, depends not only on the problem but also on the method used to solve it. When the problem is solved by the use of a computer-aided design software, one may learn a variety of skills relating to interfacing with and using that software package, including the form in which input data are required, the default assumptions and their alteration, the alternative forms or parameters in terms of which the results can be expressed, debugging errors in entered data, and many others. These are not the same skills that are learned via paper-and-pencil methods of solution, and therefore the two types of methods are not alternative routes to the acquisition of the same skills. Leaving the choice of method type to the student is to imply that the instructor has no definitive skill goals in mind.

In the following, this problem will be attacked using a variety of methods, with the focus being on the method rather than on the problem or its solution.

Method 2: Use of a Standardized Routine

A time-honored method of problem solving in engineering, sometimes called "handbook engineering," is to locate an established algorithm, such as a formula into which the given parameter values can be substituted. This method essentially replaces the higher-level cognitive tasks that the solution of the problem may otherwise require, with an established routine that can be followed almost mechanically, and takes care of the procedurization stage of problem solving. As with any off-the-shelf solution, if one can find a suitable formula and ensure its applicability, the method minimizes the time required for solution (not counting the time required for finding the right formula!). Many other types of standardized routines for problem solving are also commonplace in engineering; examples include curves, nomograms and graphical aids like Smith chart; tables of measured or computed data such as those used for

Method 2: Solution by Substitution into a Formula



S-Parameters with respect to a reference impedance Z_0:

$$S_{11} = \frac{R_3(R_1 + R_2) + (R_1 - Z_0)(R_2 + Z_0)}{R_3(R_1 + R_2 + 2Z_0) + (R_1 + Z_0)(R_2 + Z_0)}$$

$$S_{21} = \frac{2R_3 Z_0}{(R_1 + R_3 + Z_0)(R_2 + R_3 + Z_0) - R_3^2}$$

S_{22} and S_{12} determined by the same expressions after interchanging all 1's and 2's.

Given $R_1 = 20\,\Omega$, $R_2 = 10\,\Omega$, $R_3 = 120\,\Omega$, and $Z_0 = 50\,\Omega$,

$$[S] = \frac{1}{99}\begin{bmatrix} 9 & 60 \\ 60 & 4 \end{bmatrix}.$$

filter design; computer software for computation or simulation such as that employed in circuit simulation; tables of exact, approximate, or curve-fitted formulas such as used for planar transmission line design; and other experiential or empirical results and formulas. The purpose of such aids is to eliminate trial-and-error, guess work, and discretion; reduce the uncertainty, the risk of committing an error, and cognitive workload; and make the work efficient with respect to the use of time and effort.

The asymmetric-T network problem described earlier can also be solved by this method, as outlined in the sidebar "Method 2: Solution by Substitution into a Formula," if the relevant formula can be located in a reference work. Since the problem then requires minimal problem solving skills, and little more than numerical substitution of values, the problem may be a trivial exercise with a low pedagogical value. In engineering practice, however, the use of ready-made formulas is a perfectly valid, genuine, and useful problem solving method. Indeed, experts routinely rely on such chunks of knowledge to expedite the solutions, and employ them as a building block to reduce the mental workload when constructing solutions to more complex problems.

Method 3: Forward Logical Deduction

Conceptually the most transparent method of solving a problem is one in which one starts with the known information, including assertions, assumptions, and supplied details, and deduces its logical consequences. The resulting conclusions then become part of the known information, and one then proceeds to look for their logical consequences with the help of known theorems and results. In this manner, one marches forward towards the desired solution. Clearly, to ensure that one is proceeding in the right direction, one must keep an eye on the goal, and attempt to draw those conclusions that move the set of known information towards the direction of the desired destination; for this purpose the process may require strategies such as means-end analysis, difference-reduction, or elements of backwards reasoning, which is another method listed in Table 1.

The deductive method employed to deduce the consequences will usually include not only logical relationships like the transitive law, but also mathematical identities, definitions of variables, and domain-specific relationships. Solution of a problem by falling back on the fundamental definitions of terms may often not be the most expedient method of solving it, but it useful when there is uncertainty about the validity of short-cuts. The validity of the deduced results rests on that of the set of definitions and relationships employed, so that it can easily be checked or verified. If the results can be deduced solely from mathematical logic, fundamental principles, and definitions, the method has the accolade of being abinitio (from first principles).

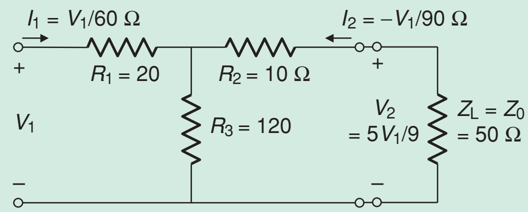
The method is illustrated in the sidebar "Method 3: Solution by Forward Logical Deduction," where the solution of the example problem starts with the given information, and successively employs circuit analysis techniques, and the basic definitions of power waves and scattering parameters, to arrive at the desired results.

Method 4: Backward Reasoning

This method is the reverse of the forward logical deduction method. Here, one starts with the desired goal, and tries to reduce it to subgoals, alternative goals, or antecedents, by analyzing the goal to determine what is needed to reach it. As in the forward deduction method, the analysis is based on the known assertions, definitions, assumptions, theorems, results, and given information. Each identified need then becomes the

Method 3: Solution by Forward Logical Deduction

Step 1. From Excitation Source to Port Voltages and Currents. For the given network excited with a source producing the voltage V_1 at the input port, and with the output port terminated in the given reference impedance Z_0, the resulting port voltages and currents, expressed in terms of the input port voltage V_1, are as marked in the figure on the right:



$$V_2 = \frac{5}{9} V_1, \qquad I_1 = \frac{1}{60\,\Omega} V_1, \qquad \text{and} \qquad I_2 = -\frac{1}{90\,\Omega} V_1.$$

Step 2. From Port Voltages/Currents to Incident and Reflected Waves. By definition, incident and reflected waves a_i and b_i at the ith port are expressed in terms of the port voltages V_1 and V_2, and port currents I_1 and I_2, as

$$a_i = \frac{V_i}{2\sqrt{Z_o}} + \frac{\sqrt{Z_o} I_i}{2} \qquad \text{and} \qquad b_i = \frac{V_i}{2\sqrt{Z_o}} - \frac{\sqrt{Z_o} I_i}{2} \qquad \text{for} \quad i = 1,\ 2$$

so that

$$a_1 = \frac{11}{6} \frac{V_1}{2\sqrt{Z_0}}, \quad b_1 = \frac{1}{6} \frac{V_1}{2\sqrt{Z_0}}, \quad b_2 = \frac{10}{9} \frac{V_1}{2\sqrt{Z_0}}, \qquad \text{and} \qquad a_2 = 0.$$

Step 3. From Incident/Reflected Waves to Port Parameters. The S-parameters are defined in terms of the incident and reflected waves are

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0} \qquad \text{for} \qquad i, j, k = 1,\ 2; \quad k \neq j$$

so that $S_{11} = 1/11$ and $S_{21} = 20/33$.

Step 4. Reuse of First Three Steps. Following a similar procedure with the input port terminated in Z_0, and the output port excited with a voltage source, the waves are

$$a_2 = \frac{198}{103} \frac{V_2}{2\sqrt{Z_0}}, \quad b_2 = \frac{8}{103} \frac{V_2}{2\sqrt{Z_0}}, \quad b_1 = \frac{120}{103} \frac{V_2}{2\sqrt{Z_0}}, \quad \text{and} \quad a_1 = 0$$

so that $S_{22} = 4/99$, and $S_{12} = 20/33$.

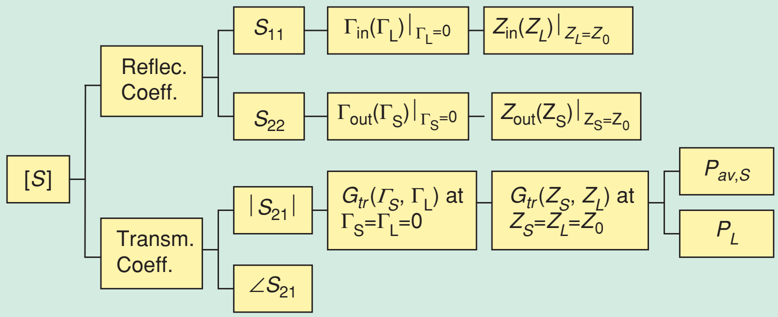
next goal, and it is analyzed further to determine what is needed to attain it. This process of working successively backwards towards the known information is continued until one reaches a set of goals whose attainment, starting from the available information, is known.

Working backwards may be a suitable strategy when the desired goal or end point is unique and known; such is the case, for example, in mathematical problems that require a proof of a given assertion, and in the reductio ad absurdum (proof by contradiction) method. The primary strength of the method is that it is efficient, because it quickly allows elimination of the unproductive paths that do not lead to the desired destination. Indeed, it is most efficient when there are only a few options (or hopefully just one) that can lead to the destination. Some backward reasoning is employed within many other problem solving methods to help select between alternatives and discard some of them. The most important limitations of this method are that the endpoint must be known in advance; as a result, the method cannot be used for exploring unexpected results.

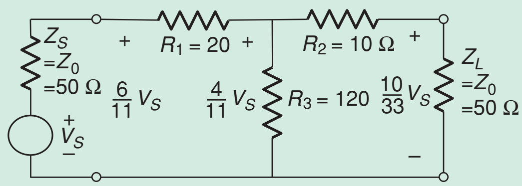
In the example problem at hand, the goal of determining [S] parameters is successively replaced by the goal of determining antecedent quantities, as shown by the flow-diagram (or tree) in the sidebar "Method 4: Solution by Backward Reasoning." The process is continued until the quantities reached are identified as known.

Method 4: Solution by Backward Reasoning

Starting with the goal of finding [S] and working backwards, the reflection coefficients S_{11} and S_{22} can be found from driving point impedances at the two respective ports (under Z_0 terminations at the other ports), while the transmission coefficient S_{21} (equal to S_{12} by reciprocity) is real and positive due to a purely resistive circuit, and is the square-root of the transducer power gain, also under port terminations of Z_0. The original goal is thus successively replaced by the goals of finding some driving point impedances and the available and load powers, as indicated in the following flowchart, each of which can be deduced immediately by observation.



With the appropriate source and loads connected,



$$Z_{in}(Z_L)|_{Z_L=Z_0} = 60\ \Omega \qquad \text{and} \qquad Z_{out}(Z_S)|_{Z_S=Z_0} = \frac{1030}{19}\ \Omega,$$

while the power level is

$$P_L = \frac{1}{50}\left(\frac{10}{33}V_s\right)^2 \qquad \text{while} \qquad P_{av,S} = \frac{1}{4\times 50}(V_s)^2.$$

Therefore,

$$S_{11} = \Gamma_{in}(\Gamma_L)|_{\Gamma L=0} = \left.\frac{Z_{in}(Z_L)-Z_0}{Z_{in}(Z_L)+Z_0}\right|_{ZL=Z0} = \frac{1}{11} \qquad \text{and} \qquad S_{22} = \Gamma_{out}(\Gamma_S)|_{\Gamma S=0} = \left.\frac{Z_{out}(Z_S)-Z_0}{Z_{out}(Z_S)+Z_0}\right|_{ZS=Z0} = \frac{4}{99},$$

while

$$S_{21} = \sqrt{G_{tr}(\Gamma_S, \Gamma_L)|_{\Gamma S=\Gamma L=0}} = \left.\frac{P_L}{P_{av,S}}\right|_{ZS=ZL=0} = \frac{20}{33}.$$

Method 6: Partitioning of the Problem

One of the most powerful methods of problem solving is to partition the problem into smaller subproblems, which can be solved separately, either in parallel or sequentially, followed by a synthesis of the subproblem solutions to arrive at the complete problem solution. The basis of partitioning can be physical (e.g., structural, spatial or temporal separation of system elements), or informational (e.g., stages in a cascaded flow of information). The solution of each subproblem may require the use of one of the other methods of solution. The method is particularly useful in problems having a high complexity, or where the subproblems are decoupled and independently (even if sequentially) solvable. Its principal drawback is the overhead effort required in partitioning and subsequent re-synthesis of components, particularly when the results of the subproblems are not already in a suitable or mutually-compatible format. When it is possible to partition a problem in more than one way, the choice between the alternatives can be based not only on the number and simplicity of resulting subproblems, but also on the ease of partitioning and re-synthesis.

In the example problem, the given asymmetric T-network can be partitioned in a variety of ways, for example by 1) subdividing R_1 into two series elements; 2) viewing R_3 as a parallel combination of two resistors, which reduces the T-network into a cascade of two L-networks; and 3) separating each resistor into a separate network, leading to a cascade of three two-port networks, each with only a single resistive element. The solution shown in the sidebar "Method 6: Solution by Partitioning of the Problem" takes advantage of the availability of [S] parameters for single resistive elements networks, but suffers from the drawback that the [S] parameters of the three networks require further transformation to [T] matrix before they can be recombined.

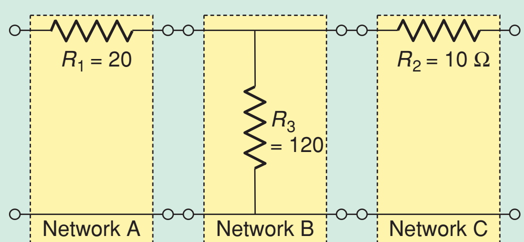
Method 9: Transformation of Goal

Often, the most efficient way to solve a problem is to solve a different problem than the one at hand, and then employ the results of that substitute problem to solve the original problem. The method is effective if the replacement problem is already known, easier, amenable to a known approach, or solvable by one of the other methods. This method is very commonly used, and many instances of this approach will already be known to the students; for example, to prove an assertion, we disprove its complement; to solve for current in a circuit, we analyze it in terms of voltages and only at the end transform back to current; and we simplify the computation of survival probabilities in reliability problem by focusing on the failed rather than the surviving components.

The crux of the method lies in finding the different problem 1) that is easier to solve than the original one and 2) whose solution enables the original

Method 6: Solution by Partitioning of the Problem

One possible partitioning of the given asymmetric-T network is as a cascade of three one-port networks, each containing only a single resistor in series or parallel.



The scattering matrices of the three subnetworks A, B, and C, defined with respect to the reference impedance of $Z_0 = 50\,\Omega$, follow immediately from textbook formulas and are:

$$S_A = \frac{1}{12}\begin{bmatrix} 2 & 10 \\ 10 & 2 \end{bmatrix}, \quad S_B = \frac{1}{29}\begin{bmatrix} -5 & 24 \\ 24 & -5 \end{bmatrix}, \quad \text{and} \quad S_C = \frac{1}{11}\begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix},$$

which can be transformed into transmission (scattering chain) matrices as

$$T_A = \frac{1}{5}\begin{bmatrix} 4 & 1 \\ -1 & 6 \end{bmatrix}, \quad T_B = \frac{1}{24}\begin{bmatrix} 19 & -5 \\ 5 & 29 \end{bmatrix}, \quad \text{and} \quad T_C = \frac{1}{10}\begin{bmatrix} 9 & 1 \\ -1 & 11 \end{bmatrix}.$$

Upon multiplication, the transmission matrix for the composite network is

$$T_A T_B T_C = \frac{1}{60}\begin{bmatrix} 36 & 9 \\ -4 & 99 \end{bmatrix},$$

which can be transformed back to scattering matrix to obtain

$$[S] = \frac{1}{99}\begin{bmatrix} 9 & 60 \\ 60 & 4 \end{bmatrix}.$$

problem to be solved. It may be possible to identify such replacement problems by altering the goal of the problem, adopting a different viewpoint, focusing on facets different from those apparent in a straightforward approach, interchanging dependent and independent parameters, transforming some variables, or by re-expressing the problem in an alternate form which may be more direct, more familiar, easier to work with, or might suggest other connections. The change of goals will require a search, for example by trial-and-error, although there are established algorithms in the field of artificial intelligence that carry out transformation of goal in one of several ways:

- by deletion of redundant subgoals, so as to reduce the complexity, or minimize diversion to better focus on the remaining goals
- by addition of implicit subgoals, so as to provide an intermediate stepping stone or pathway towards problem solution, and breakup a larger problem into shorter, less complex problems
- by clarifying a constraint or condition which may be useful as a building block in visualizing, planning, or synthesizing the complete solution.

In the example problem, the replacement of scattering matrix [S] by the impedance matrix [Z] suggests itself as a promising approach because of familiarity, ease of circuit analysis in terms of voltages and currents, and direct solution by observation. Indeed, the majority of the work incurred in this method lies in transforming the results back to the original problem, as shown in the sidebar "Method 9: Solution by Transformation of Goal."

Method 5: Search for Pattern

A pattern is a composition, arrangement, or configuration that has some characteristic feature, and is therefore distinctive or familiar. The method takes advantage of the remarkable human ability to discern a pattern that is displayed, for example, when we pick out a familiar face in a crowd, locate a drawing or geometrical figure that is camouflaged by surrounding drawings, determine the next term of a series given the first few terms, or arrange geometrical figures in an order with some succession of transformations. Symmetry, periodicity, and regularity are some of the more obvious and easily distinguished features, but are not essential to the existence of a pattern. Often one can discern a pattern within a problem, consisting of a segment that is a rudimentary, canonical or known problem. Alternatively, it may be possible to cast the problem in a form such that it (or its solution) can be arrived at by a transformation, or successive transformation, of such a familiar problem (or its solution). In such cases, it might be possible to employ the results for the known problem as a starting point, to which appropriate embedding or transformations are applied to deduce a solution of the original problem. This method can also be promising with problems involving a gradual or successive change in some geometrical feature, physical attribute, parameter value, or operation in the given system. Numerous engineering problems, such as those on ladder networks, cascaded or nested operations, and iterative or recursive operations, can be elegantly solved by this method. It is also used in many procedures for successive refinement of models or results. The principal

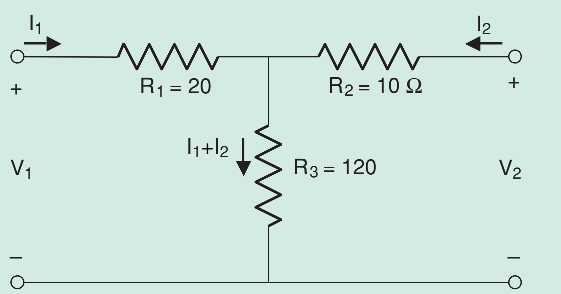
Method 9: Solution by Transformation of Goal

Instead of $[S]$ matrix, the network may be described in terms of $[Z]$ matrix, which is found directly by an application of Kirchhoff's voltage law, written in terms of port voltages V_1 and V_2, and port currents I_1 and I_2:

$$V_1 = I_1R_1 + (I_1 + I_2)R_3 \quad \Leftrightarrow \quad Z_{11}I_1 + Z_{12}I_2$$

and

$$V_2 = I_2R_2 + (I_1 + I_2)R_3 \quad \Leftrightarrow \quad Z_{21}I_1 + Z_{22}I_2$$



This immediately identifies the impedance matrix as

$$[Z] = \begin{bmatrix} R_1 + R_3 & R_3 \\ R_3 & R_2 + R_3 \end{bmatrix} = \begin{bmatrix} 140\ \Omega & 120\ \Omega \\ 120\ \Omega & 130\ \Omega \end{bmatrix}.$$

The impedance matrix can be transformed into $[S]$ matrix by standard formulae tabulated in most textbooks and handbooks,

$$[S] = \begin{bmatrix} \frac{(Z_{11}-Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}} & \frac{2Z_{12}Z_0}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}} \\ \frac{2Z_{21}Z_0}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}} & \frac{(Z_{11}+Z_0)(Z_{22}-Z_0)-Z_{12}Z_{21}}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}} \end{bmatrix}$$

to find the scattering matrix elements $S_{11} = 1/11$, $S_{22} = 4/99$, and $S_{21} = S_{12} = 20/33$.

challenge in the method arises from the difficulty of identifying a pattern, and the relationship connecting the problem to the pattern.

In the example problem, a symmetric-T network is identifiable within the given asymmetric-T network, along with a perturbation ΔR_1, as shown in the sidebar "Method 5: Solution by Pattern Search (and Extrapolation)." The scattering parameters for the symmetric attenuator are easily found by observation, and the problem then reduces to that of recovering from them the parameters for the asymmetric-T network. They can be can determined in more than one way, for example by cascading, as in the partitioning method discussed earlier. Alternatively, if only approximate values of the parameters are required, an extrapolation such as the one illustrated in the sidebar may suffice.

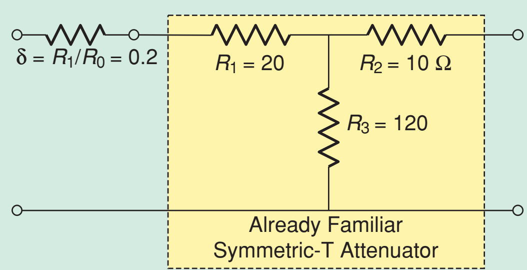
Method 10: Simplification by Approximation

A simplification can be introduced in a problem with any one of several different purposes in mind: to help reach a solution, locate a method of solution, or reduce the effort required for solution. In each case, the choice of approximations is driven by different considerations such as accuracy, generalizability, or solvability. Each of these is explained in the following.

- A problem may be simplified to make it tractable and amenable to solution, by reducing the number of variables, for example, by dropping higher order terms, or ignoring nonlinearities (if they are expected to be small) and anything else that contributes complexity out of proportion with its importance. Such approximation requires judgment, and subsequent validation. Indeed, all modeling can be viewed as a form of simplification by approximation, with the intent to leave out the unimportant. Here the criterion for selecting the approximation may be the closeness of the approximate solution to the desired solution.
- The well-known Polya method of problem solving [5] states that if you cannot solve a problem, there is a simpler problem you cannot solve; find that problem and attempt to solve it. The purpose of simplification in this case is to temporarily lift the fog of complexity so as to see the path to solution. The simplified problem is thus merely a learning vehicle, and the simplification is selected to get a foothold for climbing up the learning curve. Clearly, it is not the closeness of the approximate solution to the actual solution that is important, but only the closeness of the methods of solution, so that the method of solution employed with approximate problem can be extended to the original problem. Such approximations may be found by examining special, simple, limiting, asymptotic, or extreme cases of the problem at hand, which may lend insight into the nature of solution, and help construct a path to solution.
- An alternate purpose of simplification is to obtain an approximate solution to the problem, because an approximate solution is sufficient for the purpose at hand. Such is the case, for example, when the

Method 5: Solution by Pattern Search (and Extrapolation)

Observation: The given asymmetric-T network is composed of a symmetric-T matched attenuator, and an additional series resistor $\Delta R_1 \equiv R_1 - R_2$.



Scattering matrix for the symmetric attenuator defined with respect to the reference impedance of 50 Ω (either known, or determined more easily than that for the asymmetric network by any of the other methods) is:

$$[S_{sym}] = \begin{bmatrix} 0 & \frac{2}{3} \\ \frac{2}{3} & 0 \end{bmatrix}.$$

The exact $[S]$ matrix for the originally given asymmetric-T network can be found by using the cascading procedure as in the partitioning method.

An approximate scattering matrix of the asymmetric-T network may be found by a qualitative extrapolation to account for the effects of the perturbation caused by $\Delta R_1 \equiv R_1 - R_2 = \delta R_0$:

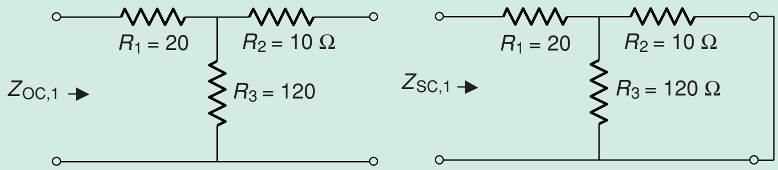
1) Introduction of a passive resistance in the R_1 arm can only decrease signal transmission in either direction, so S_{21} and S_{12} must each decrease by an amount proportional to δ.
2) Since the two ports are no longer matched, S_{11} and S_{22} must each increase above zero, and because ΔR_1 is introduced at port 1, this increase should be smaller for S_{22} than for S_{11}.

A slightly more careful accounting leads to a first-order approximation as

$$[S] \approx \begin{bmatrix} 0 + \frac{1}{2}\delta & \frac{2-\delta}{3} \\ \frac{2-\delta}{3} & 0 + \frac{1}{4}\delta \end{bmatrix}.$$

Method 10: Solution by Approximation

The given network has three specifications—the network topology, resistor values, and reference impedance—that are candidates for approximation. Consider altering one, say the reference impedance. The greatest simplification of scattering parameter determination occurs with the replacement for R_0 by the image impedances $Z_{I,1}$ and $Z_{I,2}$ at the two ports. The image impedance at a port can itself be deduced from open- and short-circuit impedances at that port, which follow from observation:



$$Z_{I,1} = \sqrt{Z_{OC,1}\,Z_{SC,1}} = \sqrt{(R_1 + R_3)[R_1 + (R_2 \| R_3)]} = 63.97\ \Omega$$

$$Z_{I,2} = \sqrt{Z_{OC,2}\,Z_{SC,2}} = \sqrt{(R_2 + R_3)[R_2 + (R_1 \| R_3)]} = 59.40\ \Omega.$$

With the choice of $Z_{I,1}$ and $Z_{I,2}$ as the reference impedances at the two respective ports, the input reflection coefficients S_{11} and S_{22} are zero, and the transmission coefficients are reduced to the square root of the power gain when the ports are terminated in the image impedances, given by

$$S_{21} = \sqrt{G_{tr}(Z_{I,1}, Z_{I,2})} = \sqrt{\frac{Z_{I,1}}{Z_{I,2}}}\left(\frac{V_2}{V_1}\right) = \sqrt{\frac{Z_{I,2}}{Z_{I,1}}}\,\frac{Z_{I,1} - R_1}{Z_{I,2} + R_2} = \frac{\sqrt{266} - \sqrt{26}}{\sqrt{247} - \sqrt{7}} \approx 0.61.$$

Consequently, the S-parameters are given by

$$[S] \approx \begin{bmatrix} 0 & 0.61 \\ 0.61 & 0 \end{bmatrix}.$$

Corrections for the errors caused by the approximation can be estimated qualitatively. Reverting back to the reference impedance of 50 Ω is equivalent to a reduction of that impedance by unequal amounts (approximately 22% and 16% respectively) at the two ports, which has two effects on the scattering parameters.

1) The values of S_{11} and S_{22} will rise due to resulting mismatches, with a smaller rise at port 2 due to the smaller reduction of reference impedance there.
2) The two terms in the transducer power gain tend to change in opposite directions, thereby cancelling most of the effect, leading to a very small reduction in S_{21} and S_{12}.

The approximate scattering parameters calculated above can therefore be treated as bounds, with the main diagonal elements being lower bounds, and off-diagonal elements being upper bounds.

input data is itself subject to large uncertainty, or because the goal of problem solving is to only make a rough estimate, or because the approximate solution is needed only to serve as the starting point in seeking a more accurate, exact, or general solution, for example by an iterative process that is expected to yield the final result. Here the criterion for selecting the approximation might be the ease of solution.

Since each of the given pieces of information in a problem is a potential candidate for approximation, multiple approximate results can be found. The method is useful when the desired result has a low sensitivity to the type of perturbations that are introduced by the approximation, or where an approximate result is useful by itself. The principal limitation of the method is that the art of clever approximation is itself a skill that has to be developed and that improves with experience.

In the asymmetric-T network example, if the network topology is fixed then the only candidates for approximation are the resistor values R_1, R_2, R_3, and the reference impedance R_0. Since the alteration of R_1 has already been demonstrated earlier,” the sidebar Method 10: Solution by Approximation” shows the results upon perturbing the reference impedance R_0 to an approximate value selected to make

(continued on page 165)

Microwave Digital Archive

One member tried to purchase previous Microwave Digital Archive Updates and were told that they were no longer available. This situation was corrected. The new publication numbers and costs are:

PUBJC18005 MTT CD-ROM 2003: US$14
PUBJC18006 MTT CD-ROM 2004: US$16
PUBJC18007 MTT CD-ROM 2005: US$20

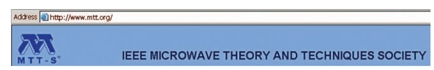
Product #PER180-CDR MTT CD-ROM 2006–2007 and 1953–2007 Index can be added to your membership for US$25.

Four members requested information on when 2006–2007 annual update and 1953–2007 DVD Index will be delivered to members. I informed them they would be delivered in May 2008. One member received the new DVD and did not realize it was for 2006 and 2007. Two members informed me that they did not receive their 2005 Microwave Digital Archive DVD and 1953–2005 Index DVD. These were sent to them.

Help Line

Much of the requested information can be found or was recently put on the MTT Society Web site, www.mtt.org.

Additional IEEE contact information is as follows: Toll free in USA & Canada +1 800 678 4333 or worldwide at +1 732 981 0060, Fax: +1 732 562 5445, Attention: member services, member-services@ieee.org. To add a new service electronically, please visit the IEEE at www.ieee.org and click on "renew." To purchase an IEEE product such as Microwave Digital Archive or updates by e-mail: customer-service@ieee.org or call the numbers listed above.



Please feel free to contact me by letter, telephone, or e-mail concerning any complaint you may have or any assistance you may need in obtaining membership services from the IEEE and MTT-S. My e-mail alias is as follows, and my mailing address and phone number may be found on the bottom of the first page of this column.

Educator's Corner *(continued from page 148)*

the S-parameter determination easiest. The approximate result may be useful by itself, serve as a basis for further refinement for example by extrapolation, or serve as bounds on the scattering parameters of the given network, since lowering R_0 can only result in 1) impedance mismatches that will raise the values of S_{11} and S_{22} and 2) reduced power transmission, which will lower S_{21} and S_{12}.

Conclusions

Awareness of various problem-solving methods is an important part of learners' metacognitive knowledge, which can be enhanced by the instructor through demonstration and discussion of the methods. The learners can better recall the methods when they are learned in the context of the subject matter content rather than in the abstract. Different learners conceptualize and approach problems differently, and presenting a variety of methods helps reach learners with diverse abilities.

References

[1] M.S. Gupta, "Framing questions to achieve higher educational objectives," *IEEE Microwave Mag.*, vol. 8, no. 4, pp. 94–105, Aug. 2007.

[2] J.D. Bransford, A.L. Brown, and R.R. Cocking, Eds., *How People Learn. Brain, Mind, Experience, and School.* Washington, D.C: National Academy Press, 1999 [Online]. Available: www.nap.edu

[3] H. Simon, "Problem Solving and Education," in *Problem Solving and Education—Issues in Teaching and Learning*, D. Tuma and F. Reif, Eds. Hillsdale, NJ: L. Erlbaum, 1980, pp. 81–94.

[4] J. Segal, S. Chipman, and R. Glaser, Eds., *Thinking and Learning Skills Vol. 1: Relating Instruction to Research* Hillsdale: L. Erlbaum, 1985.

[5] G. Polya, *How to Solve It.* Princeton, NJ: Princeton University Press, 1945.

[6] M. Levine, *Effective Problem Solving.* Englewood Cliffs, NJ: Prentice Hall, 1988.

[7] W.A. Winckelgren, *How to Solve Problems.* San Francisco, CA: Freeman, 1974.