ELEMENTARY NUMERICAL ANALYSIS

W. Allen Smith

ELEMENTARY NUMERICAL ANALYSIS

W. Allen Smith

A Reston Book
Prentice-Hall
Englewood Cliffs, New Jersey 07632

PREFACE

Elementary Numerical Analysis is geared toward college juniors, but begins easy and slowly becomes more sophisticated. The standard topics of a course in numerical analysis are presented, suitable for majors in engineering, science, and mathematics. The minimal prerequisite is elementary calculus, including exposure to series and partial derivatives. Basic concepts of Taylor series, used throughout the text, are summarized in Appendix A. For Chapters 10 and 11 it helps, but is not essential, to know a little about differential equations—about as much as might be found in many calculus books.

An effort has been made to present the concepts as clearly (and simply) as possible, with few formally stated and proven theorems. Instead of just presenting the best method for a given task, we emphasize derivations and criteria for distinguishing between competing methods. A major goal in writing this text has been to provide material that students can easily read, with ample exercises to reinforce and extend what they are learning. Earlier versions and the current version here have been class-tested for about eight years, with numerous improvements suggested by students—and a few reviewers.

A decision instructors have to make concerning numerical analysis is which computational aids are going to be emphasized: personal computers, mainframes, programmable calculators, scientific calculators, or none of these. And if a computer is used, what language is allowed. This text is designed to be used with any of these choices. Most homework exercises can be worked with the aid of an inexpensive scientific calculator, so the availability and use of a computer is not essential. However, a decided programming slant can be imparted by assigning some of the 90 Programming Projects listed throughout the sections of the text. (See "Programming projects" in the Index for page numbers.) My preference is to have students learn about important algorithms by trying them on a diverse set of examples, especially with a computer doing all the calculations. As a help in programming a calculator or computer, 18 programs—or parts of programs—are given in Appendix B. These are written in BASIC, FORTRAN, or Pascal and are unsophisticated, with little or no documentation. They are not models of good programs, just skeletons on which to build.

A different slant can be given to the course by assigning or allowing individual students to study and report on some of the 69 (nonprogramming) Project Topics listed with several of the sections. (See "Project topics" in the Index for page numbers.) In this way, many methods and related topics can be investigated that are not in this text or are only lightly touched upon.

As a quasi-practical challenge to the class, nontrivial problems can be individually assigned from Sections 3.8, 4.5, 8.4, and 10.8. Or these sections can be omitted.

Thus, this text has been written with flexibility in mind. Instructors have much freedom of choice not only as to the computational slant of the lectures, homework assignments, and possible term projects, but also which sections to cover and whether to use this text for one or two courses.

In total, there are eleven chapters plus an appendix on Taylor series and one containing sample computer programs. For the most part, the chapters are independent. Exceptions are that the concepts of error, discussed in Section 1.4, Gauss elimination, discussed in Section 2.2, and Taylor series, in Appendix A, are referred to in many of the chapters; parts of Chapters 4 and 5 depend on Chapter 3; and Chapters 9, 10, and 11 loosely form a sequence concerning differential equations.

It is possible to teach a variety of courses from this text. Four possible options are:

1. A two-course sequence. Do (nearly) all sections, using Appendix A on Taylor series together with Section 1.4 on error.
2. A single course emphasizing numerical solution of equations and curve fitting. Do Appendix A and Chapters 1, 2, 3, 5, and 7. Chapters 4 and 6 are optional. The major emphasis would be on iterative methods. My choice has been to stress nonlinear equations, leaving numerical linear algebra for a separate matrix/linear algebra course.
3. A single course emphasizing numerical calculus. Do Appendix A, Section 1.4, and Chapters 8 through 11. Chapters 6 and/or 7 could also be included, if time permits. The major emphasis would be on selecting an optimal step size along with the method.

4. A single survey course. Sections selected are at the instructor's discretion (but adhering to the interdependency of a few chapters, as stated above). Students could be assigned some sections to read on their own (without doing exercises), such as 1.1, 1.2, 1.3, Appendix A, 1.5, and 2.1. The instructor could develop Sections 1.4, 2.2, most of Chapter 3, any of the Sections 5.2 through 5.5, any of the Sections 7.2, 7.4, or 7.5, Sections 8.1 through 8.4, Section 9.3, and as much of Chapter 10 as desire and time permit.

A few facts and details. There are about 186 worked examples, 1122 exercises, and 177 true–false statements among the Problems. Answers to most of the problems are in the Answers section at the back of the book. Solutions are found in a separate *Instructor's Solutions Manual*, available from Prentice-Hall.

Numbers in brackets, especially after a name, generally refer to the References, pages 519 through 524. Important equations are numbered sequentially within a section. For instance, "7.4.6" refers to the sixth numbered equation of Section 7.4 (which is the fourth section of Chapter 7). Examples are not numbered, but those referred to later are given a letter designation, e.g., Example A. The end of a proof is denoted by a black square: ■.

I thank the students in my classes who used earlier and rough versions of this book and pointed out places where improvements might be made. Thanks also to the (anonymous to me) reviewers, whose comments were generally helpful, and to the staff at Reston Publishing Company, Inc.

W. Allen Smith

CONTENTS

Chapter 3
ROOTS OF AN EQUATION, 71

Chapter 4
ZEROES OF POLYNOMIALS, 127

Chapter 5
REAL ROOTS OF SYSTEMS OF EQUATIONS, 161

Chapter 6
DIFFERENCE OPERATORS AND THEIR APPLICATIONS, 201

Chapter 1

SOME BASIC CONCEPTS OF NUMERICAL ANALYSIS

In this chapter we show how "computational mathematics," in which approximations are usually necessary, contrasts with "exact mathematics," in which infinite precision is assumed. You could call this contrast "applied math versus pure math," but classical and modern applied math are not always identified with approximations in computing.

Apart from blunders one might make, the computer or calculator is often blamed for wrong answers. Errors *do* routinely appear when our numbers are changed into a computer's number system and also when changed back. However, both the nature of the mathematical process involved and the (finite) number system employed in the computations have much to do with errors in the output. Sources of error thus include the type of formula or procedure being used—no matter how the computations are carried out—as well as how the numbers are represented.

This chapter begins with two situations that do not involve a computing device directly, yet they have many of the qualitative features found in numerical work performed on a high-speed digital computer. In Sections 1.3 and 1.4, we discuss base two versus base ten representation and define what we mean by "error." The chapter concludes with some comments on the nature of numerical analysis and a preview of important concepts discussed in the later chapters.

Except for Section 1.4, the entire chapter can be omitted in a one-semester course.

Section 1.1
THE BLACK-BOX PROBLEM

In many problems we have a three-stage process: input (or data), generally consisting of numbers, is given; computations are performed; an output or answer is displayed. (See Figure 1.1.)

The input often comes from measurements and pertains to a mathematical model of some situation we are investigating. We think of working *within* the model, so we do not ask here whether it is valid or not. The input often contains errors and approximations, some of which we might not be aware of. However, we control the input, and can vary it if we wish.

The computations may well involve approximations, such as rounding numbers to ten decimal places, but hopefully these approximations are no worse than those in the input. If we do not look at what the particular computations are, we call the middle part a "black box." (Knowing the nature of the computations, we could call it a "white box" or "clear box.")

The output depends upon the input and the computations. In some situations

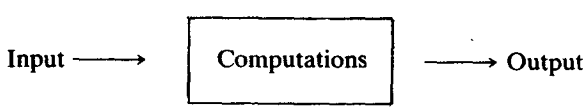

Figure 1.1 Simplified parts of a typical calculation

we seek any input that gives a prescribed output. In that case we generally do not concern ourselves with what is inside the black box but consider only the input–output relationship. For instance, the reader might think of the input as the amount of deficit spending in the national budget and the output as the rate of unemployment; the black box would consist of political and economic factors. Or the input could be the players you select from a team's roster to play in an athletic contest, such as a baseball game. The black box then involves the opponent's play, random errors, fatigue, luck, and so on, and the output consists primarily of the score, but also statistics, injuries, added experience, and the like.

EXAMPLE: We can think of the black box as a mathematical function, with the input as a domain value and the output as a range value. We are not told the function; we just obtain the output for any given (perhaps somewhat restricted) input.

Suppose we are to determine any input that results in an output of exactly 5. Without knowing anything else about the problem, what would be a reasonable first input to try? One choice might be to always start with 0 or 1, no matter what the output goal. But if the output goal is very small or very large, this is not quite as reasonable as first trying the output goal as the input. So we try 5. In other words, we assume at first that the function is the identity function.

If you are told that an input of 5 results in an output of 5.2, what would you try next? If you think the function is increasing (locally), then a smaller input would be indicated, whereas if you think the function is decreasing, then a larger input is indicated. You really need to try more inputs before distinguishing between these cases. As for the second input, it really makes no difference in the long run what you decide if nothing is known about the black box. Let us assume that the function is increasing, so we try a smaller input, say 4.

If you are now told that your input of 4 resulted in an output of 4.25, what would you try next? Of course, you may already see a pattern and can hazard a guess as to the function in the black box. Normally this is difficult, and in any case we are not trying to determine the function, only some input that will give 5. After two trials, we can work under the assumption that the function is linear. The function need not even be continuous, of course, but at first we generally try to be optimistic; if that fails, then we can later be more imaginative in our assumptions about the black box and in what we try for inputs.

If the function is linear between the inputs of 4 and 5, it is routine to calculate the desired input that gives an output of 5 as follows. Set $f(x) = a + bx$. Solving the two equations $f(5) = a + 5b = 5.2$ and $f(4) = a + 4b = 4.25$ simultaneously for a and b gives $f(x) = 0.45 + 0.95x$. Then setting $0.45 + 0.95x$ equal to 5, we get

$$x = \frac{91}{19} = 4.\overline{7894736842105263157}$$

where the bar means that the 19 digits repeat forever. It may be that the black box will accept the ratio (91/19) input form, or it may require a decimal fraction input. If it does accept the ratio, it may immediately change it to decimal form with a fixed number of places. If it is given a long string of digits in decimal form, it might use only the first few and ignore the rest. We then have the possibility, if not likelihood, of "roundoff error," which occurs because we cannot retain more than a certain number of digits, either in the input, the computations, or the output. It might be, for example, that the black box will accept and work with numbers having 50 digits, but at the end only print the six most significant digits (and keep the rest a secret). Or it could work throughout with only six digits; we could not easily distinguish between these cases.

Suppose we enter 91/19 for the input, or maybe the first 10 digits in its decimal equivalent. (Note that roundoff error in any given input is not quite so critical because we can later change the input.) If the output is now 4.9983, what should we do? One possibility is to quit, saying that 4.9983 is close enough to 5. If we did not quit, but performed more calculations similar to those in the preceding step, we might get an output of, say, 4.9997. Should we quit now? Quite possibly no input will give exactly 5, due to the method of computation (roundoff error, say) or the nature of the black-box function itself. We should not expect to exactly achieve our goal in a finite number of steps—think of adding one extra decimal place for each input attempt. So normally we stop—i.e., "truncate"—after a certain number of steps. What we then get is an approximation to the true input (if indeed it exists), and the discrepancy is called a "truncation error." Of course, what we should really do is set a goal of getting within a small fixed interval about 5. For instance, we could accept any input whose output lies in the interval (4.9999, 5.0001). This would be called a "stopping criterion," and it has a direct bearing upon the truncation, and also roundoff, error in our result. Nearly always we shall have several stopping criteria; for instance, we might also agree to stop after trying 1000 inputs if no other stopping criteria have been satisfied.

Incidentally, the function I used to generate the output numbers is $f(x) = x + 1/x$. Finding an x so that $f(x) = 5$ is equivalent to solving the quadratic equation $x^2 - 5x + 1 = 0$, whose roots are $x = (5 \pm \sqrt{21})/2$. Using decimal fraction or ratio forms (i.e., rational numbers) for inputs, we could never find either of these numbers x exactly. Hence, it is all the more reasonable to relax the requirement that the output be exactly 5.

There are many cases in which the given approach to the black-box problem, particularly the linear assumption after two trials, will fail, especially when random, variable, or discontinuous effects arise. The goal of this section was not to treat this problem exhaustively but to show why approximations and stopping criteria are often necessary.

Problems: Section 1.1

1. Suppose that in the black-box problem an input of 0 gives 7 and an input of 5 gives 9. What would be the best input to try under the linear assumption if the output goal is 10?

2. Same as problem 1, except that an input of 2 gives 5 and an input of 3 gives 2.
3. True or false?: If, in a black-box situation where the goal output is 2, an input of 2 gave 1 and an input of 1 gave 3, then 1.5 would be a good input to try next.
4. (General Case) If the goal for an output is g, an input of x_1 gives an output of y_1, and an input of x_2 gives an output of $y_2 \neq y_1$, what would be a good trial input for the third attempt?
5. Assume that a black box requires an ordered pair of numbers as input, as if it were a real-valued function $f(x,y)$.
   (*a*) Under the linear assumption, how many trials are needed before an intelligent guess can be made for the next?
   (*b*) Suppose that our output goal is 10 and we have tried three different inputs. (0, 1) gave 5, (1, 0) gave 15, and (1, 1) gave 12. If we next want to try $(2,y)$, what is a good choice for y under the linear assumption?
6. (General Case) Assume that a black box requires an ordered pair of numbers as input, as if it were a real-valued function $f(x,y)$. Suppose that our output goal is g and so far three inputs (x_i,y_i) gave different outputs z_i, $i = 1, 2, 3$. If the fourth input is to have the form $(x,0)$, what is a good choice for x under the linear assumption?
7. Ronald started with a number, raised x to this power, integrated with respect to x from 0 to 1, subtracted his original number, raised e^x to this power, took a derivative with respect to x and evaluated it at $x = 1$, took a reciprocal, added 2, added his original number, took the natural logarithm, added 1, then rounded off to five decimal places to get 0.50000.

   Estimate what number Ronald started with by applying the black-box procedure, as follows:
   (*a*) Derive the outcome associated with an input of 0.
   (*b*) Derive the outcome associated with an input of 1.862.
   (*c*) Use the formula from the answer to problem 4, or otherwise derive a good input (estimate). You need not determine the associated output.

Section 1.2
SOLVING QUADRATIC EQUATIONS

In this section we see how "significance error" can greatly distort the output from relatively simple formulas when equal or nearly equal quantities are subtracted. Standardization of a problem, in this case a polynomial, will also be treated.

Elimination of Trivial Cases

The problem we wish to consider is finding the roots of the quadratic equation

$$ax^2 + bx + c = 0 \tag{1.2.1}$$

We could consider this to be a "white-box" problem, in which x stands for the input and zero is the goal output. We would try an x and see how close $ax^2 + bx$

$+\ c$ is to zero, remembering that there can be roundoff error in the computations. We shall not treat the problem this way but instead use (quadratic) formulas for x.

If it happens that $a = 0$, then equation 1.2.1 is at most linear (that is, $bx + c = 0$) and is easy to solve. If the coefficient a is negative, we would multiply equation 1.2.1 by -1 (which does not change the roots), to get an equation with leading coefficient positive. So we shall assume that the coefficient of x^2 is positive. If $b = 0$, the roots are simply $\pm(-c/a)^{1/2}$. If $c = 0$, equation 1.2.1 factors to $x(ax + b) = 0$ and has roots $x = 0$ and $x = -b/a$. Henceforth, we shall assume that $abc \neq 0$. One consequence of this is that $x = 0$ cannot be a root.

The Quadratic Formula(s)

By completing the square in equation 1.2.1 and then taking square roots and solving for x, we find that

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{1.2.2}$$

and

$$x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{1.2.3}$$

are roots of equation 1.2.1. The numbers x_1 and x_2 are perhaps equal, perhaps complex conjugates, or perhaps real and unequal. In exact mathematics these are *the* answers and can be checked by substitution into the left of equation 1.2.1 to give (exactly) zero.

However, when we actually use these formulas we generally have to approximate the numbers involved, especially when extracting square roots. This type of approximation involves "roundoff error" and is directly related to the number of digits we are forced (by calculator or mathematical table) or choose to retain in any particular stage in the computations.

For this subsection we use "four-digit arithmetic"; that is, we carry at most four digits in each number. (These are called *significant* digits, and the intent is to use "exponential" or "scientific" notation for the numbers, with powers of 10 not counted in the four digits we carry. This notation will be further explained in the next section.) For instance, 1/3 would become .3333, 617.0853 would become 617.1, and .00345629 would become 0.003456. Leading zeroes, especially after a decimal point, do not count as part of the four digits, but inside zeroes do; thus, 6.00737 would become 6.007. Of course, to shorten a long number to four digits, we round off in a customary fashion: a digit 0 through 4 for the fifth digit means that the first four digits as given are retained; otherwise, we increase the fourth digit by 1 and drop succeeding digits. (Other rounding schemes could also be used.)

A word of caution: using four-digit arithmetic, (0.3789)(0.2614) − 0.0990 would be calculated as 0.00004, not 0.0000446, the exact result. We would keep only four significant digits, not seven, in the product of the first two numbers; then three of those four are subtracted away, so we get only one significant digit in the result.

This rounding of numbers to four significant digits introduces roundoff error, but as we shall show, this can be minor compared to the effect of subtracting nearly equal numbers, which eliminates most of the significant digits. If it is preferred, 6, 10, 12, or more digits can be kept in each number. We can get significance error (loss of significant digits) no matter how many digits are retained. In other words, no matter how many digits are retained for each number (short of an infinite number), there will be quadratic equations for which equations 1.2.2 and 1.2.3 will not give good approximations to the exact roots.

EXAMPLE: We consider solving the equation

$$x^2 - 110x + 1 = 0$$

It is known that one root is 0.009092, to four significant digits. (This might be obtained by considering the problem to be a black-box problem, for example, and happens to be what results when x^2 is omitted in the equation—which is one way to start the guessing process.) Let us see whether equation 1.2.2 or 1.2.3 gives us this root. Remembering that we can keep at most four significant digits in any number, we would calculate the so-called discriminant to be

$$b^2 - 4ac = (-110)^2 - 4(1)(1) = 12{,}100 - 4 = 12{,}100$$

The last number comes from 12,096 rounded (up) to four digits. Using equations 1.2.2 and 1.2.3, we calculate possible roots to be

$$x_1 = \frac{110 + 110}{2} = 110$$

and

$$x_2 = \frac{110 - 110}{2} = 0$$

Of course, zero cannot be an exact root, and it can be a very misleading approximation to the root 0.009092. (In Section 1.4 we shall discuss how to measure the closeness of an approximation—in other words, how to measure the error.)

If we keep more than four significant digits, the subtraction to zero in calculating x_2 disappears, and we can get closer to the correct root. However, for each fixed number of digits you decide to keep in each number, someone else can give a quadratic polynomial that will lead to the same qualitative situation. The label "significance error" applies when the output of a process has far fewer significant digits than the input. In our case, x_2 has essentially zero significant digits instead of four.

Alternative Formulas for the Roots

When significance error occurs, it will do so in only one of the two quadratic formulas, 1.2.2 or 1.2.3. When b is positive, it will occur in x_1, and when b is negative (as in the preceding example), it will occur in x_2. The best approach is to use whichever formula that does not introduce significance error to get one root. We can then use this answer in a new formula to get the second root.

We shall label the root obtained without significance error x_3. If b is positive, x_3 is calculated from equation 1.2.3, and if b is negative, x_3 is calculated from equation 1.2.2. These two cases can be written as the one equation

$$x_3 = -(\text{sign } b)\frac{|b| + \sqrt{b^2 - 4ac}}{2a} \qquad (1.2.4)$$

where sign b is $+1$ for b positive and -1 for b negative, and can be written as $b/|b| = |b|/b$. (When $b = 0$, there is no need for an alternative formula, but if completeness is desired, we can define sign $0 = 1$.)

There will be no appreciable significance error in calculating x_3 because we *add* $|b|$ to the square root, never subtract. Except for minimal roundoff error, x_3 will agree with one of the roots. But what about the other root? To determine the second root, we use the fact that the product of the roots is c/a. The derivation is as follows:

If x_3 and x_4 are roots, then $a(x - x_3)(x - x_4) = 0$ should be the original equation 1.2.1. Expanding this, we get

$$ax^2 - a(x_3 + x_4)x + ax_3x_4 = 0$$

Comparing this with $ax^2 + bx + c = 0$, we see that $ax_3x_4 = c$, so that $x_3x_4 = c/a$. Knowing x_3, we get x_4 from

$$x_4 = \frac{c}{ax_3} \qquad (1.2.5)$$

Of course, it is also true that $-a(x_3 + x_4) = b$, but to use $x_4 = -x_3 - b/a$ might introduce significance error itself if x_3 is sufficiently close to $-b/a$ (which occurs when ac is small compared to b^2).

For the preceding example we would calculate x_3 to be (approximately) 110, which is x_1. Then, from equation 1.2.5, $x_4 = 1/110 = 0.0\overline{09}$, as compared to the exact smaller root, 0.00909165 . . . .

Checking and the Lack of Standardization

The obvious way to check possible roots (in the black-box sense) is to substitute into the left side of equation 1.2.1 to see how close we come to zero. The word "close" is not precise enough for mathematics, so this common-sense approach can run into trouble, as we shall see.