JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

NUMERICAL INSTABILITY

[1]Vijayalakshmi Menon. R., [2]Santhosh P.K.

[1]Assistant Professor of Mathematics, [2]Associate Professor of Mathematics,
Department of Applied Science,
Govt. Engineering College, Kozhikode, Kerala, India

Abstract: In this paper, the concept of numerical instability is studied and thereby, the qualities required to obtain accurate result for a given problem using computers are being pointed out.

AMS Subject Classification Code: 65L20

Keywords: Numerical instability, exponent range, normalization, significant digits, relative error, inherent instability, induced instability, algorithm.

INTRODUCTION

Numerical calculation is common and necessary in all scientific fields. The high speed computing machine has made possible to find the solution of scientific and engineering problems of great complexity. So the qualities required to obtain accurate results for a given problem using computers bears much significance. In the processing of these qualities governing accuracy, the topic of numerical instability finds great prominence.

This paper is divided into three sections. Section 1, which is the basic section, deals with COMPUTER ARITHMETIC. In this section, we have a look onto how the numbers are stored in a computer. Section 2, which is the main content of this paper, deals with the different types of instability which arise during numerical computations. Finally, in section 3, we come to a conclusion regarding the qualities necessary for obtaining more accurate solutions.

I. COMPUTER ARITHMETIC [1]

1.1. FIXED-POINT FORM AND FLOATING-POINT FORM

The numbers in the computer word can be stored in two forms:

1. Fixed-point form
2. Floating-point form

In a fixed-point form, a $'t'$ digit number is assumed to have its decimal point at the left-hand end of the word. This implies that all numbers are assumed to be less than one in magnitude. The fixed-point number with base β and $'t'$ digits word length may be written as

$$\pm \sum_{k=1}^{t} \alpha_k \beta^{-k}, \text{ where } 0 \le \alpha_k \le \beta$$

To avoid the difficulty of keeping every number less than one in magnitude during computation, most computers use floating-point representation for real numbers.

A FLOATING-POINT NUMBER is characterized by 4 parameters-the base β, the number of digits $'t'$ and the exponent range (m, M). It is usually represented in the form

$$0.d_1 d_2 ... d_t \quad X \quad \beta^e$$

where $d_1, d_2 ..., d_t$ are integers and satisfy $0 \le d_i \le \beta$ and the exponent $'e'$ is such that $m \le e \le M$.

The fractional part $.d_1 d_2 ... d_t$ is called the MANTISSA and it lies between $+1$ and -1.

A non-zero floating-point number is said to be in NORMAL FORM if the value of mantissa lies in the interval $\left[-1,-\frac{.1}{\beta}\right]$ or in the interval $\left[\frac{.1}{\beta},1\right]$.

1.2. SIGNIFICANT DIGITS & RELATIVE ERROR [2]

When a number $'x'$ is written in the normalized floating-point form with $'t'$ digits in base β, we say that the number has $'t'$ SIGNIFICANT DIGITS. The leading digit $d_1 (cf 1.1)$ is called the most significant digit.

A number x^* is an approximation to x to $'t'$ significant digits if $\frac{|x-x^*|}{|x|} \le \frac{1}{2}\beta^{1-t}$, corresponding to the base β; and $\frac{1}{2}\beta^{1-t}$ is called the RELATIVE ERROR in $'x'$.

For example, $x^* = 0.3$ is an approximation to $x = 1/3$ to one significant digit with respect to the base $\beta = 10$.

These are just the contents of the basic section. Now, we move on to the main content of this paper.

II. NUMERICAL INSTABILITY [1]

Every arithmetic operation performed during computation, gives rise to some error, which may grow or decay in subsequent calculations. In some cases, the errors may grow so large as to make the computed result totally redundant. We call such a procedure NUMERICALLY UNSTABLE.

Numerical instability can be divided into two:
1. Inherent instability
2. Induced instability

2.1 INHERENT INSTABILITY

Inherent instability is the instability which arises due to the ill-conditionedness of the problem. It is the property of the problem itself. We cannot avoid inherent instability by changing the method of solution. A significant example for inherent instability is the WILKINSON's PROBLEM of finding the zeros of a polynomial. The polynomial

$$P_{20}(x) = (x-1)(x-2) \ \ldots \ (x-20)$$
$$= x^{20} - 210x^{19} + \ \ldots \ + 20!$$

has the zeros 1,2, … , 20.

Let the coefficient of x^{19} be changed from -210 to $-(210 + 2^{-23})$. This is a very small absolute change. Most computers neglect this small change which occurs after 23 binary bits. If the solution of the new equation is now computed, we find that the smaller roots are obtained with good accuracy, while roots of larger magnitude are changed by a large amount. The largest change occurs in the roots 16 and 17. They are now obtained as the complex pair $16.73 \ \ldots \ \pm i \ \ 2.81 \ \ldots$, whose magnitude is 17 approximately. This is an ample change and is due to the ill-conditionedness of the polynomial. This type of instability which arises due to the ill-conditionedness of the problem is referred to as inherent instability.

2.2 INDUCED INSTABILITY

Induced instability is the instability which arises mainly due to the wrong choice of the method of solution. Induced instability can be avoided by a suitable change of the method of solution.

For example, suppose we want to evaluate the integral

$$I_n = \int_0^1 \frac{x^n}{x+6} dx, \ \ n = 1,2, \ \ \ldots \ \ ,10.$$

The above integral can be evaluated using the recurrence relation:

$$I_n = \frac{1}{n} - 6I_{n-1}, \ \ n = 1,2, \ \ \ldots \ \ ,10. \quad \rightarrow (A)$$

where $I_0 = \int_0^1 \frac{1}{x+6} dx = \ \log\left(\frac{7}{6}\right) = 0.15415$

Using the recurrence relation (A), we get:

$$I_1 \approx 0.0751, \ \ I_2 \approx 0.0494, \ \ I_3 \approx 0.03693,$$
$$I_4 \approx 0.02842, \ \ I_5 \approx 0.02948, \ \ I_6 \approx -0.01021,$$

$I_7 \approx 0.20412, \quad I_8 \approx -1.09972, \quad I_9 \approx 6.70943,$
$I_{10} \approx -40.15658$

But the exact value $I_{10} = \int_0^1 \frac{x^{10}}{x+6} dx \quad = 0.01449$

This explosion has occurred because of the induced instability.

The problem can be solved by changing the recurrence relation (A) as

$$I_{n-1} = \frac{1}{6}\left(\frac{1}{n} - I_n\right), n = 10,9, \quad \ldots \quad ,1 \quad \rightarrow \text{(B)}$$

Since I_n decreases as n increases, we may choose $I_{10} = 0.$ Now, using the recurrence relation (B), we get

$I_9 \approx 0.01666, \quad I_8 \approx 0.01574, \quad I_7 \approx 0.01821,$
$I_6 \approx 0.02077, \quad I_5 \approx 0.02432, \quad I_4 \approx 0.02928,$
$I_3 \approx 0.03679, \quad I_2 \approx 0.04942, \quad I_1 \approx 0.0751,$
$I_0 \approx 0.15415.$

Now the exact value of I_0 is 0.15415, which shows that the problem is well- conditioned and accurate solutions can be obtained using the recurrence relation(B). This type of instability which can be avoided by a suitable change of the method of solution is called induced instability.

III. CONCLUSION

In section 2.2., we have seen that there may be more than one method to solve the same problem. Even in the computation of average of two real numbers, say $'a'$ and $'b'$, using the four digit arithmetic, it is found that the formula $C = a + \frac{b-a}{2}$ gives more accurate result than the formula $C = \frac{a+b}{2}$. So, a numerical method can be defined to be a mathematical formula for finding the solution of a given problem and we should choose the method which suits the given problem best. Once the method has been decided, we must describe a complete set of computational steps to be followed to obtain the solution. This description is called an ALGORITHM. The algorithm tells the computer where to start, what operations are to be carried out, and when to stop. The accuracy of the solution depends on the choice of the method, designing of the algorithm and computer execution.

REFERENCES

[1] JAIN M.K., 'NUMERICAL ANALYSIS FOR SCIENTISTS AND ENGINEERS', 'S.B.W. Publishers, Delhi, 1971
[2] SASTRY S.S., 'INTRODUCTORY METHODS OF NUMERICAL ANALYSIS', Fifth Edition, PHI Learning Private Limited, Delhi