

## Chapter 01.03

### Sources of Error

*After reading this chapter, you should be able to:*

1. *know that there are two inherent sources of error in numerical methods – round-off and truncation error,*
2. *recognize the sources of round-off and truncation error, and*
3. *know the difference between round-off and truncation error.*

Error in solving an engineering or science problem can arise due to several factors. First, the error may be in the modeling technique. A mathematical model may be based on using assumptions that are not acceptable. For example, one may assume that the drag force on a car is proportional to the velocity of the car, but actually it is proportional to the square of the velocity of the car. This itself can create huge errors in determining the performance of the car, no matter how accurate the numerical methods you may use are. Second, errors may arise from mistakes in programs themselves or in the measurement of physical quantities. But, **in applications of numerical methods itself, the two errors we need to focus on are**

1. **Round off error**
2. **Truncation error.**

**Q:** What is round off error?

**A:** A computer can only represent a number approximately. For example, a number like  $\frac{1}{3}$  may be represented as 0.333333 on a PC. Then the round off error in this case is  $\frac{1}{3} - 0.333333 = 0.000000\bar{3}$ . Then there are other numbers that cannot be represented exactly. For example,  $\pi$  and  $\sqrt{2}$  are numbers that need to be approximated in computer calculations.

**Q:** What problems can be created by round off errors?

**A:** Twenty-eight Americans were killed on February 25, 1991. An Iraqi Scud hit the Army barracks in Dhahran, Saudi Arabia. The patriot defense system had failed to track and intercept the Scud. What was the cause for this failure?

The Patriot defense system consists of an electronic detection device called the range gate. It calculates the area in the air space where it should look for a Scud. To find out where it

should aim next, it calculates the velocity of the Scud and the last time the radar detected the Scud. Time is saved in a register that has 24 bits length. Since the internal clock of the system is measured for every one-tenth of a second,  $1/10$  is expressed in a 24 bit-register as 0.00011001100110011001100. However, this is not an exact representation. In fact, it would need infinite numbers of bits to represent  $1/10$  exactly. So, the error in the representation in decimal format is



**Figure 1** Patriot missile (Courtesy of the US Armed Forces, <http://www.redstone.army.mil/history/archives/patriot/patriot.html>)

$$\frac{1}{10} - (0 \times 2^{-1} + 0 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4} + \dots + 1 \times 2^{-22} + 0 \times 2^{-23} + 0 \times 2^{-24})$$

$$= 9.537 \times 10^{-8}$$

The battery was on for 100 consecutive hours, hence causing an inaccuracy of

$$= 9.537 \times 10^{-8} \frac{\text{s}}{0.1\text{s}} \times 100 \text{ hr} \times \frac{3600\text{s}}{1\text{hr}}$$

$$= 0.3433\text{s}$$

The shift calculated in the range gate due to 0.3433s was calculated as 687m. For the Patriot missile defense system, the target is considered out of range if the shift was going to more than 137m.

**Q:** What is truncation error?

**A:** Truncation error is defined as the error caused by truncating a mathematical procedure.

For example, the Maclaurin series for  $e^x$  is given as

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

This series has an infinite number of terms but when using this series to calculate  $e^x$ , only a finite number of terms can be used. For example, if one uses three terms to calculate  $e^x$ , then

$$e^x \approx 1 + x + \frac{x^2}{2!}.$$

the truncation error for such an approximation is

$$\begin{aligned}\text{Truncation error} &= e^x - \left(1 + x + \frac{x^2}{2!}\right) \\ &= \frac{x^3}{3!} + \frac{x^4}{4!} + \dots\end{aligned}$$

But, how can truncation error be controlled in this example? We can use the concept of relative approximate error to see how many terms need to be considered. Assume that one is calculating  $e^{1.2}$  using the Maclaurin series, then

$$e^{1.2} = 1 + 1.2 + \frac{1.2^2}{2!} + \frac{1.2^3}{3!} + \dots$$

Let us assume one wants the absolute relative approximate error to be less than 1%. In Table 1, we show the value of  $e^{1.2}$ , approximate error and absolute relative approximate error as a function of the number of terms,  $n$ .

$n$	$e^{1.2}$	$E_a$	$ \epsilon_a \%$
1	1	-	-
2	2.2	1.2	54.546
3	2.92	0.72	24.658
4	3.208	0.288	8.9776
5	3.2944	0.0864	2.6226
6	3.3151	0.020736	0.62550

Using 6 terms of the series yields a  $|\epsilon_a| < 1\%$ .

**Q:** Can you give me other examples of truncation error?

**A:** In many textbooks, the Maclaurin series is used as an example to illustrate truncation error. This may lead you to believe that truncation errors are just chopping a part of the series. However, truncation error can take place in other mathematical procedures as well. For example to find the derivative of a function, we define

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

But since we cannot use  $\Delta x \rightarrow 0$ , we have to use a finite value of  $\Delta x$ , to give

$$f'(x) \approx \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

So the truncation error is caused by choosing a finite value of  $\Delta x$  as opposed to a  $\Delta x \rightarrow 0$ .

For example, in finding  $f'(3)$  for  $f(x) = x^2$ , we have the exact value calculated as follows.

$$f(x) = x^2$$

From the definition of the derivative of a function,

$$\begin{aligned}f'(x) &= \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{(x + \Delta x)^2 - (x)^2}{\Delta x}\end{aligned}$$

$$\begin{aligned}
 &= \lim_{\Delta x \rightarrow 0} \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x} \\
 &= \lim_{\Delta x \rightarrow 0} (2x + \Delta x) \\
 &= 2x
 \end{aligned}$$

This is the same expression you would have obtained by directly using the formula from your differential calculus class

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

By this formula for

$$\begin{aligned}
 f(x) &= x^2 \\
 f'(x) &= 2x
 \end{aligned}$$

The exact value of  $f'(3)$  is

$$\begin{aligned}
 f'(3) &= 2 \times 3 \\
 &= 6
 \end{aligned}$$

If we now choose  $\Delta x = 0.2$ , we get

$$\begin{aligned}
 f'(3) &= \frac{f(3 + 0.2) - f(3)}{0.2} \\
 &= \frac{f(3.2) - f(3)}{0.2} \\
 &= \frac{3.2^2 - 3^2}{0.2} \\
 &= \frac{10.24 - 9}{0.2} \\
 &= \frac{1.24}{0.2} \\
 &= 6.2
 \end{aligned}$$

We purposefully chose a simple function  $f(x) = x^2$  with value of  $x = 2$  and  $\Delta x = 0.2$  because we wanted to have no round-off error in our calculations so that the truncation error can be isolated. The truncation error in this example is

$$6 - 6.2 = -0.2.$$

Can you reduce the truncate error by choosing a smaller  $\Delta x$  ?

Another example of truncation error is the numerical integration of a function,

$$I = \int_a^b f(x) dx$$

Exact calculations require us to calculate the area under the curve by adding the area of the rectangles as shown in Figure 2. However, exact calculations requires an infinite number of such rectangles. Since we cannot choose an infinite number of rectangles, we will have truncation error.

For example, to find

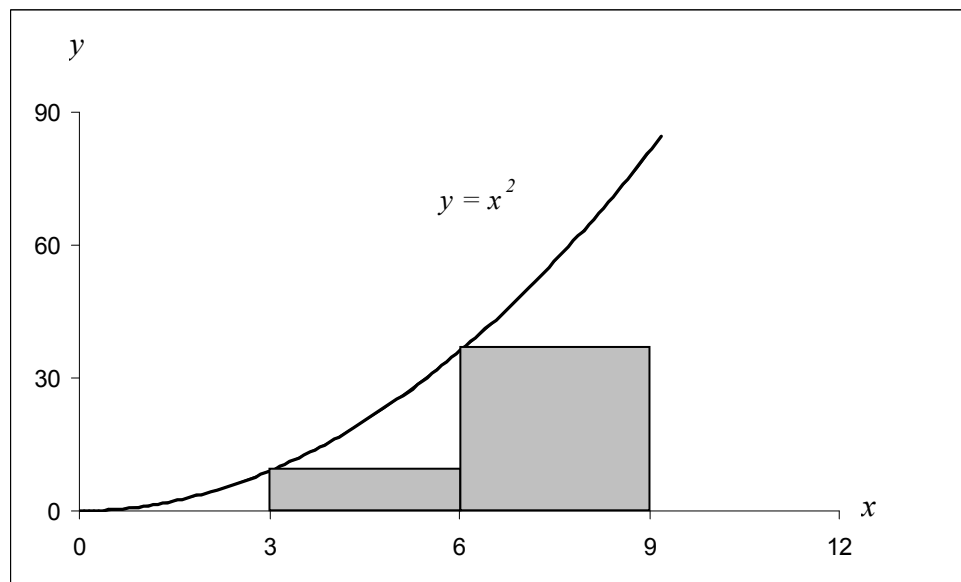
$$\int_3^9 x^2 dx,$$

we have the exact value as

$$\begin{aligned}\int_3^9 x^2 dx &= \left[ \frac{x^3}{3} \right]_3^9 \\ &= \left[ \frac{9^3 - 3^3}{3} \right] \\ &= 234\end{aligned}$$

If we now choose to use two rectangles of equal width to approximate the area (see Figure 2) under the curve, the approximate value of the integral

$$\begin{aligned}\int_3^9 x^2 dx &= (x^2) \Big|_{x=3} (6-3) + (x^2) \Big|_{x=6} (9-6) \\ &= (3^2)3 + (6^2)3 \\ &= 27 + 108 \\ &= 135\end{aligned}$$

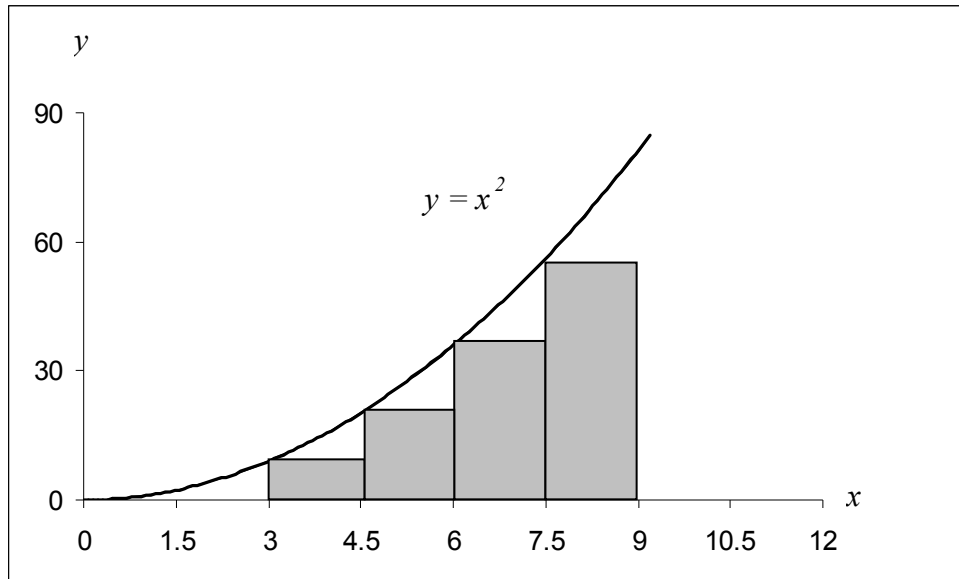


**Figure 2** Plot of  $y = x^2$  showing the approximate area under the curve from  $x = 3$  to  $x = 9$  using two rectangles.

Again, we purposefully chose a simple example because we wanted to have no round off error in our calculations. This makes the obtained error purely truncation. The truncation error is

$$234 - 135 = 99$$

Can you reduce the truncation error by choosing more rectangles as given in Figure 3? What is the truncation error?



**Figure 3** Plot of  $y = x^2$  showing the approximate area under the curve from  $x = 3$  to  $x = 9$  using four rectangles.

### References

“Patriot Missile Defense – Software Problem Led to System Failure at Dhahran, Saudi Arabia”, GAO Report, General Accounting Office, Washington DC, February 4, 1992.

---

### INTRODUCTION, APPROXIMATION AND ERRORS

---

Topic	Sources of error
Summary	Textbook notes on sources of error
Major	General Engineering
Authors	Autar Kaw
Date	December 23, 2009
Web Site	<a href="http://numericalmethods.eng.usf.edu">http://numericalmethods.eng.usf.edu</a>

---