

Optimization and Numerical Methods

Syllabus:-

1- Introduction

- *Introduction*
- *Measuring Errors*
- *Sources of Error*
- *Propagation of Errors*

2- Non Linear Equations

- *Bisection Method*
- *False- Position Method*
- *Simple Fixed point Method*
- *Newton-Raphson Method*
- *Secant Method*

3- Simultaneous Linear Equations

- *Gauss Elimination Method*
- *Gauss Jordan Elimination Method*
- *Cramer's Rule Method*
- *Gauss-Seidel Method*
- *SOR Method*
- *Jacobi Method*

4- Interpolation

- *Direct Method*
- *Newton's Quadratic Method*
- *Lagrange Method*
- *Rational Method*
- *Spline Method*

5- Regression

- *Linear Regression*
- *Polynomial Regression*
- *Nonlinear Regression*

6- Integration

- *Trapezoidal Rule*
- *Simpson's Rule*

7- Ordinary Differential Equations

- *Taylor's Series method*
- *Euler's Method*
- *Improved Euler's Method*
- *Runge-Kutta 2nd Method*
- *Runge-Kutta 4th Method*

8- Optimization

Introduction to Numerical Methods

After reading this chapter, you should be able to:

- 1- understand the need for numerical methods, and*
- 2- go through the stages (mathematical modeling, solving and implementation) of solving a particular physical problem.*

Mathematical models are an integral part in solving engineering problems. Many times, these mathematical models are derived from engineering and science principles, while at other times the models may be obtained from experimental data.

Mathematical models generally result in need of using mathematical procedures that include but are not limited to

- (A) differentiation,*
- (B) nonlinear equations,*
- (C) simultaneous linear equations,*
- (D) curve fitting by interpolation or regression,*
- (E) integration, and*
- (F) differential equations.*

These mathematical procedures may be suitable to be solved exactly as you must have experienced in the series of calculus courses you have taken, but in most cases, the procedures need to be solved approximately using numerical methods. Let us see an example of such a need from a real-life physical problem.

To make the fulcrum (Figure 1) of a bascule bridge, a long hollow steel shaft called the trunnion is shrink fit into a steel hub. The resulting steel trunnion-hub assembly is then shrink fit into the girder of the bridge.

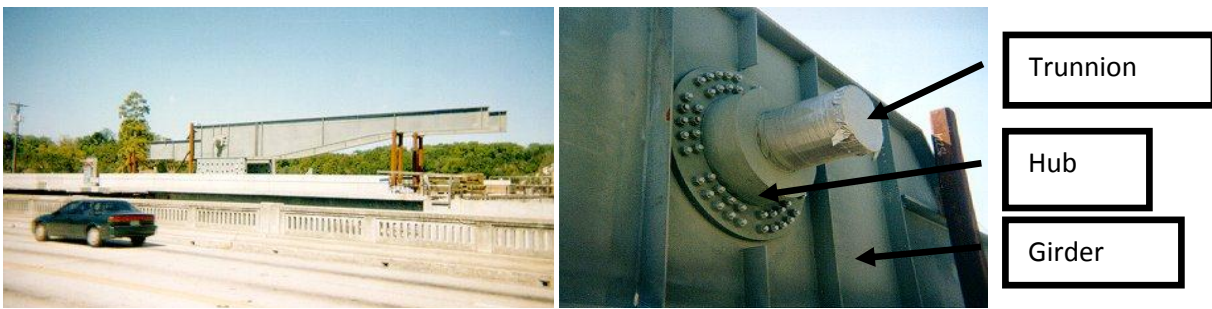


Figure 1 Trunnion-Hub-Girder (THG) assembly.

This is done by first immersing the trunnion in a cold medium such as a dry ice/alcohol mixture. After the trunnion reaches the steady state temperature of the cold medium, the trunnion outer diameter contracts. The trunnion is taken out of the medium and slid through the hole of the hub (Figure 2).

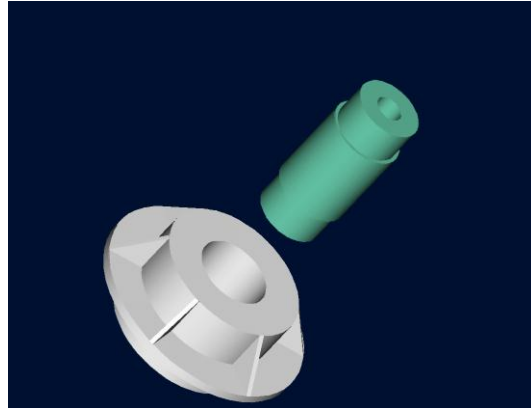


Figure 2 *Trunnion slid through the hub after contracting*

When the trunnion heats up, it expands and creates an interference fit with the hub. In 1995, on one of the bridges in Florida, this assembly procedure did not work as designed. Before the trunnion could be inserted fully into the hub, the trunnion got stuck. Luckily, the trunnion was taken out before it got stuck permanently. Otherwise, a new trunnion and hub would be needed to be ordered at a cost of \$50,000. Coupled with construction delays, the total loss could have been more than a hundred thousand dollars.

Why did the trunnion get stuck? *This was because the trunnion had not contracted enough to slide through the hole. Can you find out why?*

A hollow trunnion of outside diameter 12.363" is to be fitted in a hub of inner diameter 12.358". The trunnion was put in dry ice/alcohol mixture (temperature of the fluid - dry ice/alcohol mixture is -108°F) to contract the trunnion so that it can be slid through the hole of the hub. To slide the trunnion without sticking, a diametrical clearance of at least 0.01" is required between the trunnion and the hub. Assuming the room temperature is 80°F , is immersing the trunnion in dry-ice/alcohol mixture a correct decision?

To calculate the contraction in the diameter of the trunnion, the thermal expansion coefficient at room temperature is used. In that case the reduction ΔD in the outer diameter of the trunnion is

$$\Delta D = D\alpha\Delta T \quad (1)$$

Where

D = outer diameter of the trunnion,

α = coefficient of thermal expansion coefficient at room temperature, and

ΔT = change in temperature,

Given

$$D = 12.363''$$

$$\alpha = 6.47 \times 10^{-6} \text{ in/in/}^\circ\text{F at } 80^\circ\text{F}$$

$$\Delta T = T_{\text{fluid}} - T_{\text{room}}$$

$$= -108 - 80$$

$$= -188^\circ\text{F}$$

Where

T_{fluid} = temperature of dry-ice/alcohol mixture

T_{room} = room temperature

The reduction in the outer diameter of the trunnion is given by

$$\Delta D = (12.363)(6.47 \times 10^{-6})(-188)$$

$$= -0.01504''$$

So the trunnion is predicted to reduce in diameter by 0.01504". But, is this enough reduction in diameter? As per specifications, the trunnion needs to contract by

= trunnion outside diameter - hub inner diameter + diametric clearance

$$= 12.363 - 12.358 + 0.01$$

$$= 0.015''$$

So according to his calculations, immersing the steel trunnion in dry-ice/alcohol mixture gives the desired contraction of greater than 0.015" as the predicted contraction is 0.01504". But, when the steel trunnion was put in the hub, it got stuck.

Why did this happen? Was our mathematical model adequate for this problem or did we create a mathematical error?

As shown in Figure 3 and Table 1, the thermal expansion coefficient of steel decreases with temperature and is not constant over the range of temperature the trunnion goes through. Hence, Equation (1) would overestimate the thermal contraction.

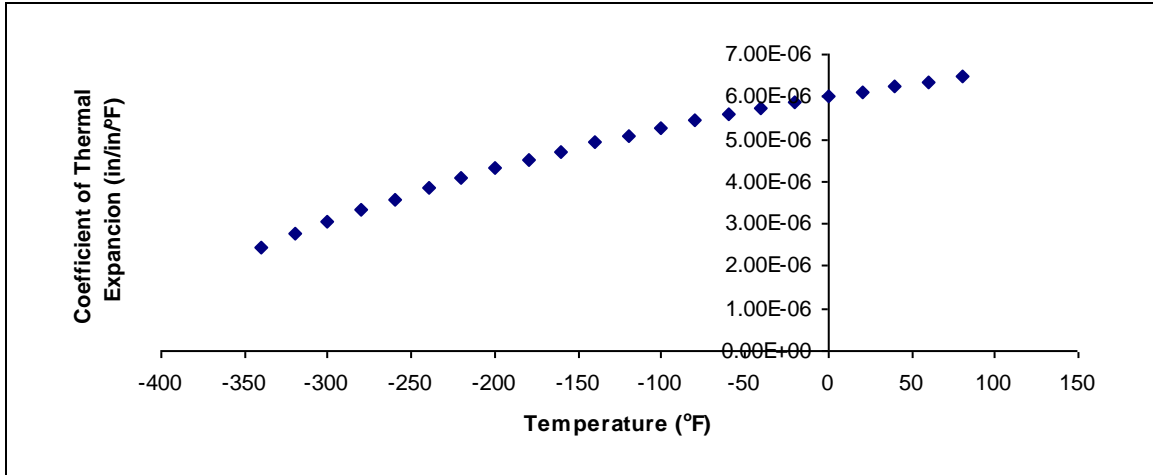


Figure 3 Varying thermal expansion coefficient as a function of temperature for cast steel.

The contraction in the diameter of the trunnion for which the thermal expansion coefficient varies as a function of temperature is given by

$$\Delta D = D \int_{T_{room}}^{T_{fluid}} \alpha dT \tag{2}$$

So one needs to curve fit the data to find the coefficient of thermal expansion as a function of temperature. This is done by regression where we best fit a curve through the data given in Table 1. In this case, we may fit a second order polynomial

$$\alpha = a_0 + a_1 \times T + a_2 \times T^2 \tag{3}$$

Table 1 Instantaneous thermal expansion coefficient as a function of temperature.

Temperature	Instantaneous Thermal Expansion
°F	μin/in/°F
80	6.47
60	6.36
40	6.24
20	6.12
0	6.00
-20	5.86
-40	5.72
-60	5.58

-80	5.43
-100	5.28
-120	5.09
-140	4.91
-160	4.72
-180	4.52
-200	4.30
-220	4.08
-240	3.83
-260	3.58
-280	3.33
-300	3.07
-320	2.76
-340	2.45

The values of the coefficients in the above Equation (3) will be found by polynomial regression (we will learn how to do this later). At this point we are just going to give you these values and they are

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 6.0150 \times 10^{-6} \\ 6.1946 \times 10^{-9} \\ -1.2278 \times 10^{-11} \end{bmatrix}$$

To give the polynomial regression model (Figure 4) as

$$\begin{aligned} \alpha &= a_0 + a_1 T + a_2 T^2 \\ &= 6.0150 \times 10^{-6} + 6.1946 \times 10^{-9} T - 1.2278 \times 10^{-11} T^2 \end{aligned}$$

Knowing the values of a_0 , a_1 and a_2 , we can then find the contraction in the trunnion diameter as

$$\begin{aligned} \Delta D &= D \int_{T_{room}}^{T_{fluid}} (a_0 + a_1 T + a_2 T^2) dT \\ &= D \left[a_0 (T_{fluid} - T_{room}) + a_1 \frac{(T_{fluid}^2 - T_{room}^2)}{2} + a_2 \frac{(T_{fluid}^3 - T_{room}^3)}{3} \right] \end{aligned} \quad (4)$$

Which gives

$$\begin{aligned} \Delta D &= 12.363 \left[\begin{aligned} &6.0150 \times 10^{-6} \times (-108 - 80) + 6.1946 \times 10^{-9} \frac{((-108)^2 - (80)^2)}{2} \\ &- 1.2278 \times 10^{-11} \frac{((-108)^3 - (80)^3)}{3} \end{aligned} \right] \\ &= -0.013689" \end{aligned}$$

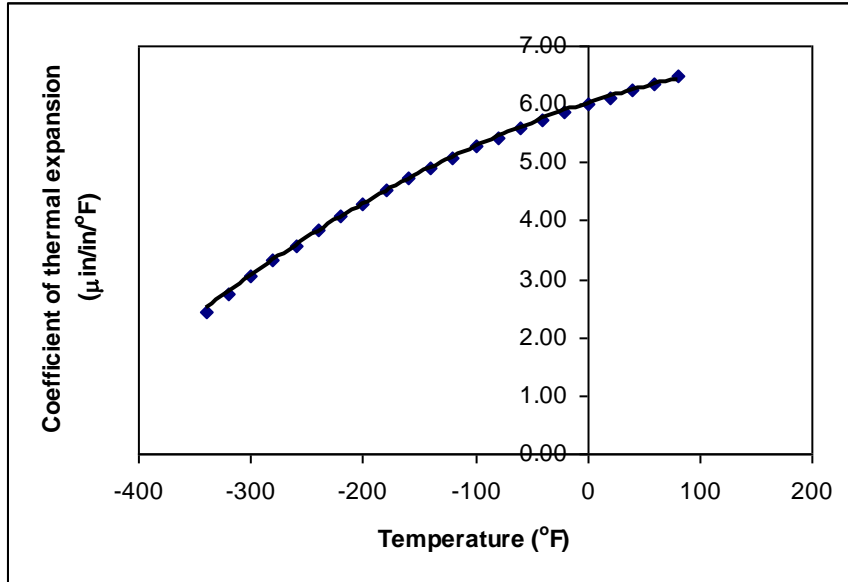


Figure 4 Second order polynomial regression model for coefficient of thermal expansion as a function of temperature.

What do we find here? The contraction in the trunnion is not enough to meet the required specification of 0.015".

❖ ***So what is the solution to the problem?***

One solution is to immerse the trunnion in liquid nitrogen which has a boiling point of -321°F as opposed to the dry-ice/alcohol temperature of -108°F.

$$\Delta D = -0.0244"$$

❖ ***Revisiting steps to solve a problem***

1) ***Problem Statement:*** Trunnion got stuck in the hub.

2) ***Modeling:*** Developed a new model

$$\Delta D = D \int_{T_a}^{T_c} \alpha(T) dT$$

3) ***Solution:*** a) Used trapezoidal rule OR

b) Used regression and integration.

4) ***Implementation:*** Cool the trunnion in liquid nitrogen

Measuring Errors

After reading this chapter, you should be able to:

- 1. find the true and relative true error,*
- 2. find the approximate and relative approximate error,*
- 3. relate the absolute relative approximate error to the number of significant digits at least correct in your answers, and*
- 4. know the concept of significant digits.*

In any numerical analysis, errors will arise during the calculations. To be able to deal with the issue of errors, we need to

- (G) identify where the error is coming from, followed by*
- (H) quantifying the error, and lastly*
- (I) minimize the error as per our needs.*

In this chapter, we will concentrate on item (B), that is, how to quantify errors.

Q: What is true error?

A: True error denoted by E_t is the difference between the true value (also called the exact value) and the approximate value.

True Error = True value – Approximate value

Example 1

The derivative of a function $f(x)$ at a particular value of x can be approximately calculated by

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

of $f'(2)$ For $f(x) = 7e^{0.5x}$ and $h = 0.3$, find

- a) the approximate value of $f'(2)$*
- b) the true value of $f'(2)$*
- c) the true error for part (a)*

Solution

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

For $x = 2$ and $h = 0.3$,

$$\begin{aligned}
f'(2) &\approx \frac{f(2+0.3) - f(2)}{0.3} \\
&= \frac{f(2.3) - f(2)}{0.3} \\
&= \frac{7e^{0.5(2.3)} - 7e^{0.5(2)}}{0.3} \\
&= \frac{22.107 - 19.028}{0.3} \\
&= 10.265
\end{aligned}$$

b) The exact value of $f'(2)$ can be calculated by using our knowledge of differential calculus.

$$\begin{aligned}
f(x) &= 7e^{0.5x} \\
f'(x) &= 7 \times 0.5 \times e^{0.5x} \\
&= 3.5e^{0.5x}
\end{aligned}$$

So the true value of $f'(2)$ is

$$\begin{aligned}
f'(2) &= 3.5e^{0.5(2)} \\
&= 9.5140
\end{aligned}$$

c) True error is calculated as

$$\begin{aligned}
E_t &= \text{True value} - \text{Approximate value} \\
&= 9.5140 - 10.265 \\
&= -0.75061
\end{aligned}$$

The magnitude of true error does not show how bad the error is. This brings us to the definition of relative true error.

Q: What is relative true error?

A: Relative true error is denoted by ϵ_t and is defined as the ratio between the true error and the true value.

$$\text{Relative True Error} = \frac{\text{True Error}}{\text{True Value}}$$

Example 2

The derivative of a function $f(x)$ at a particular value of x can be approximately calculated by

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

For $f(x) = 7e^{0.5x}$ and $h = 0.3$, find the relative true error at $x = 2$.

Solution

From Example 1,

$$\begin{aligned} E_t &= \text{True value} - \text{Approximate value} \\ &= 9.5140 - 10.265 \\ &= -0.75061 \end{aligned}$$

Relative true error is calculated as

$$\begin{aligned} \epsilon_t &= \frac{\text{True Error}}{\text{True Value}} \\ &= \frac{-0.75061}{9.5140} \\ &= -0.078895 \end{aligned}$$

Relative true errors are also presented as percentages. For this example,

$$\begin{aligned} \epsilon_t &= -0.0758895 \times 100\% \\ &= -7.58895\% \end{aligned}$$

Absolute relative true errors may also need to be calculated. In such cases,

$$\begin{aligned} |\epsilon_t| &= |-0.0758888| \\ &= 0.0758895 \\ &= 7.58895\% \end{aligned}$$

Q: What is approximate error?

A: *In the previous section, we discussed how to calculate true errors. Such errors are calculated only if true values are known. An example where this would be useful is when one is checking if a program is in working order and you know some examples where the true error is known. **But mostly we will not have the luxury of knowing true values as why you would want to find the approximate values if you know the true values.** So when we are solving a problem numerically, we will only have access to approximate values. We need to know how to quantify error for such cases.*

Approximate error is denoted by E_a and is defined as the difference between the present approximation and previous approximation.

$$\text{Approximate Error} = \text{Present Approximation} - \text{Previous Approximation}$$

Example 3

The derivative of a function $f(x)$ at a particular value of x can be approximately calculated by

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

For $f(x) = 7e^{0.5x}$ and at $x = 2$, find the following

- $f'(2)$ using $h = 0.3$
- $f'(2)$ using $h = 0.15$
- approximate error for the value of $f'(2)$ for part (b)

Solution

a) The approximate expression for the derivative of a function is

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

For $x = 2$ and $h = 0.3$,

$$\begin{aligned} f'(2) &\approx \frac{f(2+0.3) - f(2)}{0.3} \\ &= \frac{f(2.3) - f(2)}{0.3} \\ &= \frac{7e^{0.5(2.3)} - 7e^{0.5(2)}}{0.3} \\ &= \frac{22.107 - 19.028}{0.3} \\ &= 10.265 \end{aligned}$$

b) Repeat the procedure of part (a) with $h = 0.15$,

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

For $x = 2$ and $h = 0.15$,

$$\begin{aligned} f'(2) &\approx \frac{f(2+0.15) - f(2)}{0.15} \\ &= \frac{f(2.15) - f(2)}{0.15} \\ &= \frac{7e^{0.5(2.15)} - 7e^{0.5(2)}}{0.15} \\ &= \frac{20.50 - 19.028}{0.15} \\ &= 9.8799 \end{aligned}$$

c) So the approximate error, E_a is

$$\begin{aligned} E_a &= \text{Present Approximation} - \text{Previous Approximation} \\ &= 9.8799 - 10.265 \\ &= -0.38474 \end{aligned}$$

The magnitude of approximate error does not show how bad the error is. This brings us to the definition of relative approximate error.

Q: What is relative approximate error?

A: Relative approximate error is denoted by ϵ_a and is defined as the ratio between the approximate error and the present approximation.

$$\text{Relative Approximate Error} = \frac{\text{Approximate Error}}{\text{Present Approximation}}$$

Example 4

The derivative of a function $f(x)$ at a particular value of x can be approximately calculated by

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

For $f(x) = 7e^{0.5x}$, find the relative approximate error in calculating $f'(2)$ using values from $h = 0.3$ and $h = 0.15$.

Solution

From Example 3, the approximate value of $f'(2) = 10.263$ using $h = 0.3$ and $f'(2) = 9.8800$ using $h = 0.15$.

$$\begin{aligned} E_a &= \text{Present Approximation} - \text{Previous Approximation} \\ &= 9.8799 - 10.265 \\ &= -0.38474 \end{aligned}$$

The relative approximate error is calculated as

$$\begin{aligned} \epsilon_a &= \frac{\text{Approximate Error}}{\text{Present Approximation}} \\ &= \frac{-0.38474}{9.8799} \\ &= -0.038942 \end{aligned}$$

Relative approximate errors are also presented as percentages. For this example,

$$\begin{aligned} \epsilon_a &= -0.038942 \times 100\% \\ &= -3.8942\% \end{aligned}$$

Absolute relative approximate errors may also need to be calculated. In this example

$$\begin{aligned} |\epsilon_a| &= |-0.038942| \\ &= 0.038942 \text{ or } 3.8942\% \end{aligned}$$

Q: While solving a mathematical model using numerical methods, how can we use relative approximate errors to minimize the error?

A: In a numerical method that uses iterative methods, a user can calculate relative approximate error ϵ_a at the end of each iteration. The user may pre-specify a minimum acceptable tolerance called the pre-specified tolerance, ϵ_s . If the absolute relative approximate error ϵ_a is less than or equal to the pre-specified tolerance ϵ_s , that is, $|\epsilon_a| \leq \epsilon_s$, then the acceptable error has been reached and no more iterations would be required. Alternatively, one may pre-specify how many significant digits they would like to be correct in their answer. In that case, if one wants at least m significant digits to be correct in the answer, then you would need to have the absolute relative approximate error, $|\epsilon_a| \leq 0.5 \times 10^{2-m} \%$.

Example 5

If one chooses 6 terms of the Maclaurin series for e^x to calculate $e^{0.7}$, how many significant digits can you trust in the solution? Find your answer without knowing or using the exact answer.

Solution

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

Using 6 terms, we get the current approximation as

$$\begin{aligned} e^{0.7} &\cong 1 + 0.7 + \frac{0.7^2}{2!} + \frac{0.7^3}{3!} + \frac{0.7^4}{4!} + \frac{0.7^5}{5!} \\ &= 2.0136 \end{aligned}$$

Using 5 terms, we get the previous approximation as

$$\begin{aligned} e^{0.7} &\cong 1 + 0.7 + \frac{0.7^2}{2!} + \frac{0.7^3}{3!} + \frac{0.7^4}{4!} \\ &= 2.0122 \end{aligned}$$

The percentage absolute relative approximate error is

$$|\epsilon_a| = \left| \frac{2.0136 - 2.0122}{2.0136} \right| \times 100 \\ = 0.069527\%$$

Since $|\epsilon_a| \leq 0.5 \times 10^{2-2}\%$, at least 2 significant digits are correct in the answer of $e^{0.7} \cong 2.0136$

Q: But what do you mean by significant digits?

A: Significant digits are important in showing the truth one has in a reported number. For example, if someone asked me what the population of my county is, I would respond, "The population of the Hillsborough county area is 1 million". But if someone was going to give me a \$100 for every citizen of the county, I would have to get an exact count. That count would have been 1,079,587 in year 2003. So you can see that in my statement that the population is 1 million, that there is only one significant digit, that is, 1, and in the statement that the population is 1,079,587, there are seven significant digits. So, how do we differentiate the number of digits correct in 1,000,000 and 1,079,587? Well for that, one may use scientific notation. For our data we show

$$1,000,000 = 1 \times 10^6$$

$$1,079,587 = 1.079587 \times 10^6$$

to signify the correct number of significant digits.

Example 5

Give some examples of showing the number of significant digits.

Solution

- a) 0.0459 has three significant digits
- b) 4.590 has four significant digits
- c) 4008 has four significant digits
- d) 4008.0 has five significant digits
- e) 1.079×10^3 has four significant digits
- f) 1.0790×10^3 has five significant digits
- g) 1.07900×10^3 has six significant digits

Sources of Error

After reading this chapter, you should be able to:

5. know that there are two inherent sources of error in numerical methods – round-off and truncation error,
6. recognize the sources of round-off and truncation error, and
7. 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. **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

$\frac{1}{3} - 0.333333 = 0.00000033$. Then there are other numbers that cannot be represented exactly. For example, π and $\sqrt{2}$ are numbers that need to be approximated in computer calculations.

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

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

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

$$f'(x) = 2x$$

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 Δx ?

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 require 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

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

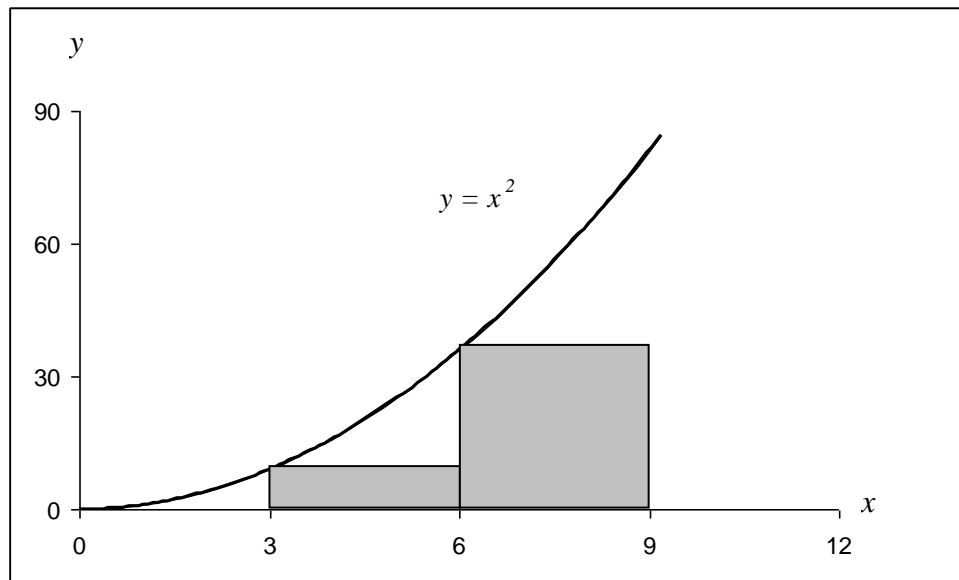


Figure 2 Plot of $y = x^2$ showing the approximate area under the curve from $x = 3$ to $x = 9$ using two rectangles.

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? What is the truncation error?

Propagation of Errors

If a calculation is made with numbers that are not exact, then the calculation itself will have an error. How do the errors in each individual number propagate through the calculations? Let's look at the concept via some examples.

Example 1

Find the bounds for the propagation error in adding two numbers. For example if one is calculating $X + Y$ where

$$X = 1.5 \pm 0.05,$$

$$Y = 3.4 \pm 0.04 .$$

Solution

By looking at the numbers, the maximum possible values of X and Y are

$$X = 1.55 \text{ and } Y = 3.44$$

Hence

$$X + Y = 1.55 + 3.44 = 4.99$$

is the maximum value of $X + Y$.

The minimum possible values of X and Y are

$$X = 1.45 \text{ and } Y = 3.36 .$$

Hence

$$\begin{aligned} X + Y &= 1.45 + 3.36 \\ &= 4.81 \end{aligned}$$

is the minimum value of $X + Y$.

Hence

$$4.81 \leq X + Y \leq 4.99.$$

What if the evaluations we are making are function evaluations instead? How do we find the value of the propagation error in such cases?

If f is a function of several variables $X_1, X_2, X_3, \dots, X_{n-1}, X_n$, then the maximum possible value of the error in f is

$$\Delta f \approx \left| \frac{\partial f}{\partial X_1} \Delta X_1 \right| + \left| \frac{\partial f}{\partial X_2} \Delta X_2 \right| + \dots + \left| \frac{\partial f}{\partial X_{n-1}} \Delta X_{n-1} \right| + \left| \frac{\partial f}{\partial X_n} \Delta X_n \right|$$

Example 2

The strain in an axial member of a square cross-section is given by

$$\epsilon = \frac{F}{h^2 E}$$

where

F = axial force in the member, N

h = length or width of the cross-section, m

E = Young's modulus, Pa

Given

$$F = 72 \pm 0.9 \text{ N}$$

$$h = 4 \pm 0.1 \text{ mm}$$

$$E = 70 \pm 1.5 \text{ GPa}$$

Find the maximum possible error in the measured strain.

Solution

$$\begin{aligned} \epsilon &= \frac{72}{(4 \times 10^{-3})^2 (70 \times 10^9)} \\ &= 64.286 \times 10^{-6} \\ &= 64.286 \mu \end{aligned}$$

$$\Delta \epsilon = \left| \frac{\partial \epsilon}{\partial F} \Delta F \right| + \left| \frac{\partial \epsilon}{\partial h} \Delta h \right| + \left| \frac{\partial \epsilon}{\partial E} \Delta E \right|$$

$$\frac{\partial \epsilon}{\partial F} = \frac{1}{h^2 E}$$

$$\frac{\partial \epsilon}{\partial h} = -\frac{2F}{h^3 E}$$

$$\begin{aligned} \frac{\partial \epsilon}{\partial E} &= -\frac{F}{h^2 E^2} \\ \Delta E &= \left| \frac{1}{h^2 E} \Delta F \right| + \left| \frac{2F}{h^3 E} \Delta h \right| + \left| \frac{F}{h^2 E^2} \Delta E \right| \\ &= \left| \frac{1}{(4 \times 10^{-3})^2 (70 \times 10^9)} \times 0.9 \right| + \left| \frac{2 \times 72}{(4 \times 10^{-3})^3 (70 \times 10^9)} \times 0.0001 \right| \\ &\quad + \left| \frac{72}{(4 \times 10^{-3})^2 (70 \times 10^9)^2} \times 1.5 \times 10^9 \right| \\ &= 8.0357 \times 10^{-7} + 3.2143 \times 10^{-6} + 1.3776 \times 10^{-6} \\ &= 5.3955 \times 10^{-6} \\ &= 5.3955 \mu \end{aligned}$$

Hence

$$\epsilon = (64.286 \mu \pm 5.3955 \mu)$$

implying that the axial strain, ϵ is between 58.8905μ and 69.6815μ

Example 3

Subtraction of numbers that are nearly equal can create unwanted inaccuracies. Using the formula for error propagation, show that this is true.

Solution

Let

$$z = x - y$$

Then

$$\begin{aligned} |\Delta z| &= \left| \frac{\partial z}{\partial x} \Delta x \right| + \left| \frac{\partial z}{\partial y} \Delta y \right| \\ &= |(1)\Delta x| + |(-1)\Delta y| \\ &= |\Delta x| + |\Delta y| \end{aligned}$$

So the absolute relative change is

$$\left| \frac{\Delta z}{z} \right| = \frac{|\Delta x| + |\Delta y|}{|x - y|}$$

As x and y become close to each other, the denominator becomes small and hence create large relative errors.

For example if

$$x = 2 \pm 0.001$$

$$y = 2.003 \pm 0.001$$

$$\frac{|\Delta z|}{|z|} = \frac{|0.001| + |0.001|}{|2 - 2.003|}$$

$$= 0.6667$$

$$= 66.67\%$$