

# ANALYSIS OF NUMERICAL ERRORS

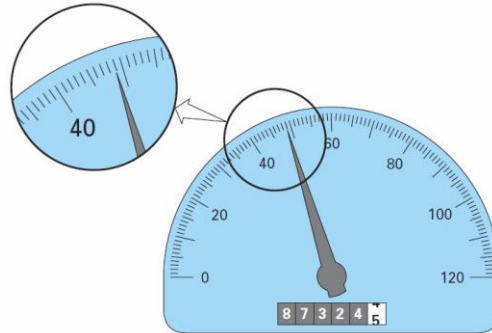
## SIGNIFICANT FIGURES

By visual inspection of the speedometer in Fig 3.1, could you tell at what speed is this car traveling?

A: between 48 and 49 and most people would claim is traveling close to 49 km/h.

**FIGURE 3.1**

An automobile speedometer and odometer illustrating the concept of a significant figure.



Let's estimate the speed to one decimal place: Some will say 48.8 others 49.9. Difficult due to the limitations of the instrument. Estimates of the third digit (or higher) must be viewed as approximations. It would be ridiculous to claim, on the basis of this speedometer, that the automobile is traveling at 48.8642138 km/h.

The concept of a significant figure, or digit, has been developed to formally designate the **reliability** of a numerical value. The significant digits of a number are those that can be used with confidence. They correspond to the number of certain digits plus **one** estimated digit. For example, the speedometer and the odometer in Fig. 3.1 yield readings of three and seven significant figures, respectively.

For the speedometer, the two certain digits are 48. It is conventional to set the estimated digit at one-half of the smallest scale division on the measurement device, therefore:

Speedometer = 48.5 (3 significant figures)

Odometer = 87,324.45 (7 significant figures)

### Zeros are not always significant

The following numbers all have four significant figures:

- 0.00001913
- 0.0001913,
- 0.001913

### Trailing zeros

Trailing zeros may create confusion, e.g. 45,300

This number may have three, four, or five significant digits, depending on whether the zeros are known with confidence.

### Scientific notation

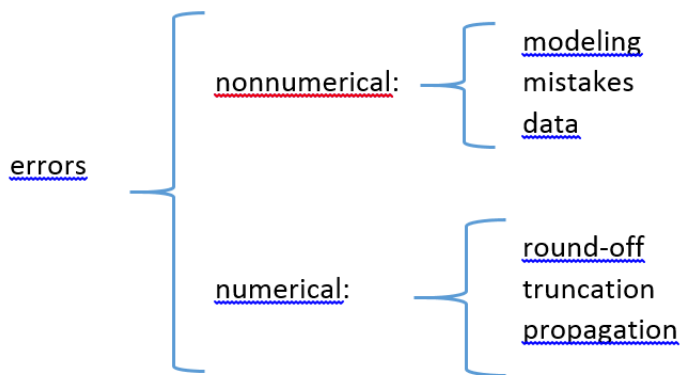
Provides a convenient way where the above number can be expressed as

- $4.53 \times 10^4$ ,
- $4.530 \times 10^4$ ,
- $4.5300 \times 10^4$

Designate that the number is known to three, four, and five significant figures, respectively.

Numerical methods yield approximate results. Therefore, developing criteria to specify how confident we are in the approximate result is necessary. One way to do this is in terms of significant figures. For example, we might decide that our approximation is acceptable if it is correct to five significant figures.

### TYPES OF ERRORS



### ROUND-OFF

- computer limitation to store most floating point numbers
- a terminating decimal fraction can be nonterminating in base 2:  
e.g.

$$0.6_{10} \longrightarrow 0.1001100110011_2$$

- some computers round the final answer others just chop off extra digits

### TRUNCATION ERROR

- results from the truncation of a series or number:

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

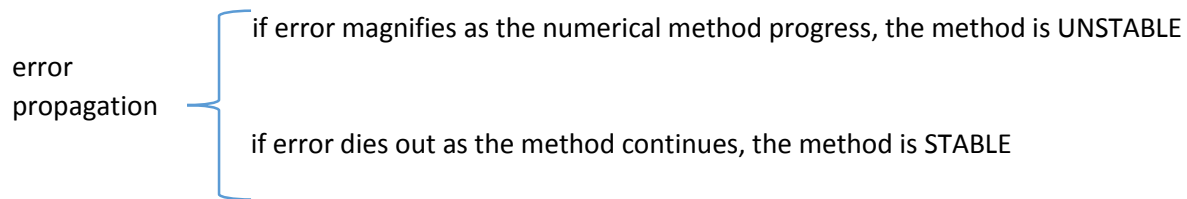
truncation

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \underbrace{\vartheta(x^4)}_{\text{error of the order of } x^4}$$

- truncation error is the difference between the truncated expression and the nontruncated expression

#### PROPAGATION

- Errors propagate in multiple operations if intermediate results are not expressed or stored with a significant number of digits
- To avoid propagation, intermediate calculations should be done with more significant digits than the number of digits of the data. Computed values should only be rounded at the end of the computation process



#### ACCURACY (es: exactitud)

- accurate values are close to the true value
- deviation from the true value

#### PRECISION (es: precisión)

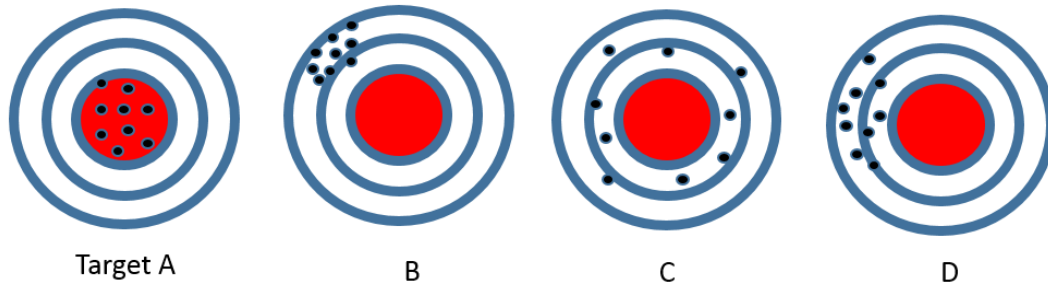
- multiple estimates close to each other

#### BIAS (es: inexactitud o sesgo)

- systematic deviation of values from the true value

#### Examples

These concepts can be illustrated graphically using an analogy from target practice. The bullet holes on each target in the figure below can be thought of as the predictions of a numerical technique, whereas the bull's-eye represents the true value. A shooter has different performance in the following targets.



Target	PRECISION	ACCURACY	BIAS
A	High	High	None (unbiased)
B	High	Low	High
C	Low	Low	None (unbiased)
D	Low	Low	Moderate



## ERROR DEFINITIONS

e.g., produced by a numerical method

$$\text{true value} = \text{approximation} + \text{error}$$

By rearranging we find that the numerical error is equal to the discrepancy between the truth and the approximation, as in

$$E_t = \text{true value} - \text{approximation}$$

The following expression takes into account the magnitude of quantities being evaluated

$$\text{True fractional relative error} = \frac{\text{true error}}{\text{true value}}$$

Also

$$\varepsilon_t = \frac{\text{true error}}{\text{true value}} 100\%$$

Eq 3.3

where  $\varepsilon_t$  is the true percent relative error

The availability of an analytical solution allow us to compute the error exactly. For many problems, we cannot obtain analytical solutions. Therefore, we cannot compute exactly the errors associated with our numerical methods. In these cases, we must settle for approximations or estimates of the errors.

In real-world applications, many times we don't not know the true answer a priori. For these situations, an alternative is to normalize the error using the best available estimate of the true value as in

$$\varepsilon_a = \frac{\text{approximate error}}{\text{approximation}} 100\%$$

Certain numerical methods use an iterative approach to compute answers. A current approximation is made on the basis of a previous approximation. This process is performed repeatedly, or iteratively, to successively compute better and better approximations. For such cases, the error is often estimated as the difference between previous and current approximations. Thus, percent relative error is determined according to

$$\varepsilon_a = \frac{\text{current approximation} - \text{previous approximation}}{\text{current approximation}} 100\%$$

Eq 3.5

Often, when performing computations, we may not be concerned with the sign of the error, but we are interested in whether the percent absolute value is lower than a prespecified percent tolerance  $\varepsilon_t$ . For such cases, the computation is repeated until

$$|\varepsilon_a| < \varepsilon_s$$

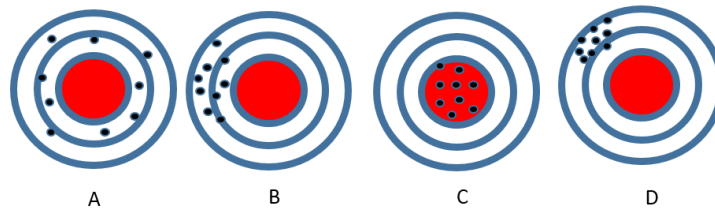
error tolerance

Eq 3.6

It is also important to relate these errors to the number of significant figures in the approximation. It can be shown (Scarborough, 1966) that if the following criterion is met, we can be assured that the result is correct to at least n significant figures.

$$\varepsilon_s = (0.5 \times 10^{2-n})\% \quad \text{Eq 3.7}$$

## EXAMPLES



Target	PRECISION	ACCURACY	BIAS
A			
B			
C			
D			

### EXAMPLE 3.1 Calculation of Errors

**Problem Statement.** Suppose that you have the task of measuring the lengths of a bridge and a rivet and come up with 9999 and 9 cm, respectively. If the true values are 10,000 and 10 cm, respectively, compute (a) the true error and (b) the true percent relative error for each case.

**Solution.**

(a) The error for measuring the bridge is [Eq. (3.2)]

$$E_t = 10,000 - 9999 = 1 \text{ cm}$$

and for the rivet it is

$$E_t = 10 - 9 = 1 \text{ cm}$$

(b) The percent relative error for the bridge is [Eq. (3.3)]

$$\varepsilon_t = \frac{1}{10,000} 100\% = 0.01\%$$

and for the rivet it is

$$\varepsilon_t = \frac{1}{10} 100\% = 10\%$$

Thus, although both measurements have an error of 1 cm, the relative error for the rivet is much greater. We would conclude that we have done an adequate job of measuring the bridge, whereas our estimate for the rivet leaves something to be desired.

### EXAMPLE 3.2 Error Estimates for Iterative Methods

**Problem Statement.** In mathematics, functions can often be represented by infinite series. For example, the exponential function can be computed using

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \quad (\text{E3.2.1})$$

Thus, as more terms are added in sequence, the approximation becomes a better and better estimate of the true value of  $e^x$ . Equation (E3.2.1) is called a *Maclaurin series expansion*.

Starting with the simplest version,  $e^x = 1$ , add terms one at a time to estimate  $e^{0.5}$ . After each new term is added, compute the true and approximate percent relative errors with Eqs. (3.3) and (3.5), respectively. Note that the true value is  $e^{0.5} = 1.648721 \dots$ . Add terms until the absolute value of the approximate error estimate  $\varepsilon_a$  falls below a prespecified error criterion  $\varepsilon_s$  conforming to three significant figures.

**Solution.** First, Eq. (3.7) can be employed to determine the error criterion that ensures a result is correct to at least three significant figures:

$$\varepsilon_s = (0.5 \times 10^{2-3})\% = 0.05\%$$

Thus, we will add terms to the series until  $\varepsilon_a$  falls below this level.

The first estimate is simply equal to Eq. (E3.2.1) with a single term. Thus, the first estimate is equal to 1. The second estimate is then generated by adding the second term, as in

$$e^x = 1 + x$$

or for  $x = 0.5$ ,

$$e^{0.5} = 1 + 0.5 = 1.5$$

This represents a true percent relative error of [Eq. (3.3)]

$$\varepsilon_t = \frac{1.648721 - 1.5}{1.648721} 100\% = 9.02\%$$

Equation (3.5) can be used to determine an approximate estimate of the error, as in

$$\varepsilon_a = \frac{1.5 - 1}{1.5} 100\% = 33.3\%$$

Because  $\varepsilon_a$  is not less than the required value of  $\varepsilon_s$ , we would continue the computation by adding another term,  $x^2/2!$ , and repeating the error calculations. The process is continued until  $\varepsilon_a < \varepsilon_s$ . The entire computation can be summarized as

Terms	Result	$\varepsilon_t$ (%)	$\varepsilon_a$ (%)
1	1	39.3	
2	1.5	9.02	33.3
3	1.625	1.44	7.69
4	1.645833333	0.175	1.27
5	1.648437500	0.0172	0.158
6	1.648697917	0.00142	0.0158

Thus, after six terms are included, the approximate error falls below  $\varepsilon_s = 0.05\%$  and the computation is terminated. However, notice that, rather than three significant figures, the result is accurate to five! This is because, for this case, both Eqs. (3.5) and (3.7) are conservative. That is, they ensure that the result is at least as good as they specify. Although, as discussed in Chap. 6, this is not always the case for Eq. (3.5), it is true most of the time.