Chapter 10: Error Analysis

Error Analysis

To complete the solution of a numerical problem, we need some estimate errors.

Source of Errors:

- **Measurement Errors**

determined by accuracy of measuring instruments and built-in bias of equipment and conditions.

For example, an instrument may be able to record values for a particular physical quantity only to the nearest one tenth (0.1) of a unit.

- Truncation Errors

due to approximations that use finite sequence of operations. E.g.,

- $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \pm \ldots$ (error equal to numerical value of omitted terms)

- Newton Root-Finding Methods, etc.

- **Roundoff Errors**

due to finite precision of computer storage.

For example,

$$0.1_{10} = 0.00011001\ldots_2$$

Thus, some decimal values do not have an exact internal numerical representation.

Can get roundoff with input values and with computed values.

Methods for Describing Errors

- **Absolute Error = Actual Value - Approx. Value**

But we usually don't know the actual value.
So use Bound on Absolute Error. E.g.,

$$x = 2.374 \pm 0.001$$

where $\varepsilon = 0.001$ is called

Absolute Error Bound

and $2.373 \le x \le 2.375$

- $$\text{Relative Error} = \frac{\text{Absolute Error}}{\text{Actual Data Value}}$$

(fractural representation or error)
Since actual value rarely known, use

$$\boxed{\begin{array}{c}\text{Relative Error} \\ \text{Bound}(\rho)\end{array} = \frac{\text{Absolute Error Bound}}{\text{Approximation Data Value}}}$$

E.g., Suppose $x = 1.500$ (approx. value)
with $\varepsilon = 0.015$. Then

$$\rho = \frac{\varepsilon}{x} = \frac{0.015}{1.500} = 0.01 \qquad or \qquad \rho = 1\%$$

and we can express the approx. value and error as

$$1.500 \pm 1\%$$

i.e., $1.500\,(1 \pm 1\%) = 1.500\,(1 \pm \rho)$

Significant Digits

A measured or calculated quantity usually has some uncertainty, which limits the number of digits that are "significant" in the representation of the quantity.

For example, $x = 3.74$ implies that x has three significant digits with some uncertainty in the last digit. In the absence of any other information about the error associated with this quantity, we can take the error to be $\pm$ 0.005; i.e.,

$$3.735 \leq x \leq 3.745$$

We determine the number of significant digits in a value by counting the number of specified digits starting at the leftmost nonzero digit to the left of the decimal point. Thus,

0.456 - has 3 significant digits

0815.0 - has 4 significant digits

<u>Propagation of Errors</u>

Consider two values x1, x2 with error bounds

$$\varepsilon_{x1}, \varepsilon_{x2}, \rho_{x1}, \rho_{x2}$$

1. Addition

$$sum = (x1 \pm \varepsilon x_1) + (x1 \pm \varepsilon_{x2})$$

$$= (x1 + x2) \pm (\varepsilon_{x1} + \varepsilon_{x2})$$

with absolute error bound:

$$\varepsilon s_{um} = \varepsilon_{x1} + \varepsilon_{x2}$$

<u>Example</u>: $sum = (4.678 \;\; 0.001) + (1.236 \pm 0.005)$

$$= 5.914 \pm 0.006$$

i.e., $5.908 \leq sum \leq 5.920$ (with four significant digits)

When two numbers are combined, the result cannot have more significant digits than either of the original numbers. E.g.,

$$sum = (15.2 \pm 0.1) + (0.010 \pm 0.003)$$

$$= 15.210 \;\; 0.102$$

But results cannot have more that three significant digits. Therefore,

$sum = 15.2 \pm 0.1$ and significant digits in second number are lost.

It may be possible to avoid this in a particular problem by rearranging the order of the calculations; i.e., accumulate small numbers before adding to larger numbers.

2. Subtraction

$diff = (x1 \pm \varepsilon_{x1}) - (x2 \pm \varepsilon_{x2})$
$\quad = (x1 - x2) \pm (\varepsilon_{x1} + \varepsilon_{x2})$

$$\boxed{\varepsilon_{diff} = \varepsilon_{x1} + \varepsilon_{x2}}$$

Absolute error bound same as addition.

<u>Example</u>:

$diff = (4.678 \pm 0.001) - (1.236 \pm 0.005)$
$\quad = 3.442 \pm 0.006$

or $\quad 3.436 \leq diff \leq 3.448$

3. Multiplication

$prod = x1(1 \pm \rho_{x1}) \cdot x2(1 \pm \rho_{x2})$
$\quad = x1 \cdot x2\ (1 \pm \rho_{x1} \pm \rho_{x2} \pm \rho_{x1}\rho_{x2})$

Assuming the error product $\rho_{x1}\rho_{x2}$ is much smaller than either $\rho_{\bar{x}1}$ or ρ_{x2}, we can neglect the product term.

Thus,

$$prod = x1 \cdot x2\ [1 \pm \rho_{x1} + \rho_{x2})]$$

and relative error bound for the product is

$$\boxed{\rho_{prod} = \rho_{x1} + \rho_{x2}}$$

with

$$\boxed{\text{Absolute Error Bound} = \frac{x1}{x2}(\rho_{div})}$$

4. Division

$$div = \frac{x1(1 \pm \rho_{x1})}{x2(1 \pm \rho_{x2})}$$

$$= \frac{x1(1 \pm \rho_{x1})}{x2}(1 \mp \rho_{x2} \pm \rho_{x2}^2 \mp \rho_{x2}^3 \pm \ldots)$$

$$= \frac{x1}{x2}(1 \pm \rho_{x1})(1 \pm \rho_{x2}) \qquad \text{(neglecting terms with power > 1)}$$

$$= \frac{x1}{x2}[1 \pm (\rho_{x1} + \rho_{x2} + \rho_{x1}\rho_{x2})]$$

Neglecting the product term, we have

$$div = \frac{x1}{x2}[1 \pm (\rho_{x1} + \rho_{x2})]$$

Relative error bound is thus the same as for multiplication:

$$\boxed{\rho_{div} = \rho_{x1} + \rho_{x2}}$$

and

$$\boxed{\text{Absolute Error Bound} = \frac{x1}{x2}(\rho_{div})}$$

<u>Example</u> - Calculating Errors

$$x1 = 3.500 \qquad \varepsilon_{x1} = 0.001$$

$$x2 = 2.70 \qquad \varepsilon_{x2} = 0.001$$

For addition or subtraction, absolute error bound is

$$\varepsilon_{x1} + \varepsilon_{x2} = 0.002 = 2 \times 10^{-3}$$

For multiplication or division, relative error bound is

$$\begin{aligned}
\rho_{x1} + \rho_{x2} &= \frac{\varepsilon_{x1}}{x1} + \frac{\varepsilon_{x2}}{x2} \\
&= \frac{10^{-3}}{3.500} + \frac{10^{-3}}{2.701} \\
&\approx (0.29 + 0.37) \times 10^{-3} \\
&\approx 0.7 \times 10^{-3} \qquad (0.07\%)
\end{aligned}$$

And corresponding absolute error bound for multiplication is

$$\begin{aligned}
x1 \cdot x2(\rho_{x2}) &= (3.500)(2.70)(0.7 \times 10^{-3}) \\
&\approx 7 \times 10^{-3}
\end{aligned}$$

(What is absolute error bound for division?)

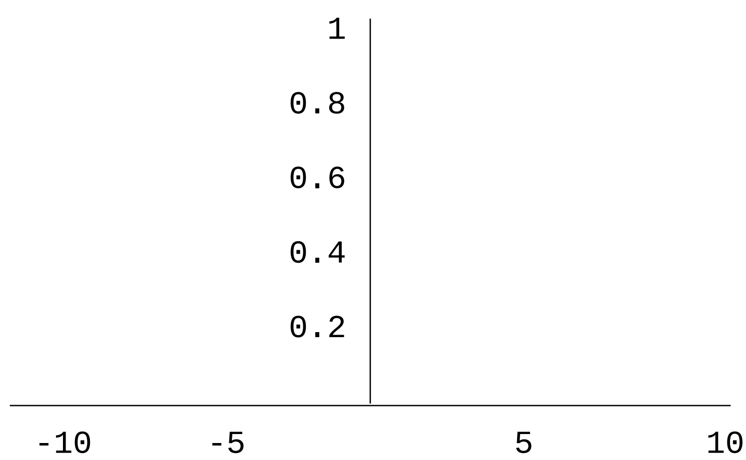
Numerical Errors in Computations

Besides roundoff and truncation errors, we can also pickup errors in our computations due to “sampling” procedures.

In numerical integration, for instance, the integration range and subinterval positions can cause significant areas to be unsampled if the function has many oscillations or sharp peaks.

The following function is an example of this:

```
Plot[Exp[-x^4],{x,-10,10}, PlotRange->All]
```



This function has a narrow peak centered on x = 0. If the integration range is not too wide, the function values near 0 will be properly samples.

```
NIntegrate[Exp[-x^4], {x, -10, 10}]
```

```
1.8128
```

But for larger integration intervals, we begin to pick up some sampling errors.

```
NIntegrate[Exp[-x^4], {x, -100, 100}]

NIntegrate::slwcon:
    Numerical integration converging too slowly;
        suspect one of thefollowing: singularity,
        oscillagory integrand, or insufficient
        WorkingPrecision.

NIntegrate::slwcon:
    Numerical integration converging too slowly;
        suspect one of the following: singularity,
        oscillatory integrand, or isufficient
        WorkingPrecision.

NIntegrate::ncvb:
    NIntegrate failed to converge to prescribed
        accuracy after 7 recursive bisections
        in x near x = 0.78125.

1.813
```

(At least Mathematica tells us about the problem.)

And for a very wide integration range, we can miss the peak altogether:

```
NIntegrate[Exp[-x^4], {x, -500, 500}]

NIntegrate::ploss
    Numerical integration stopping due to loss
        of precision. Achieved neither the
        requested PrecisionGoal no AccuracyGoal;
        suspect one of the following: highly
        oscillatory integrand or the true value
        of the integral is 0.
0.
```

We can often improve the numerical evaluation of such integrals using Monte Carlo techniques with a large number of random points.

Similar problems can occur in other types of numerical evaluations, such as computing the value of a series of terms or locating function extrema.

In many computational problems, we can reduce errors by carrying out operations sybolically as far as possible. Numerical evaluations are then performed only at the last step.

Complex Numbers - Review

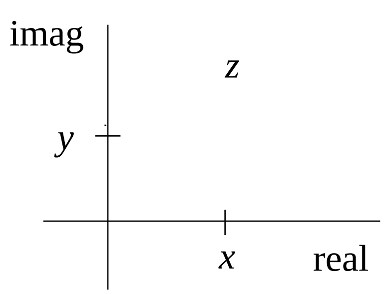
Need for complex number representations arises from solution of equation such as $x^2 + 1 = 0$ which has no real-number solution.

Thus need to introduce an extension to real numbers:

Define:

complex number = ordered pair of real numbers (with special rules for arithmetic operations)

i.e., special type of 2D vector:
$z = (x,y)$, with x, y as real numbers



x - real part of z,

y - imaginary part of z.

Complex Arithmetic

- Addition/Subtraction:

$$(x_1, y_1) \pm (x_2, y_2) = (x_1 \pm x_2, y_1 \pm y_2)$$

- Multiplication By Scalar (Real Number):

$$a(x, y) = (ax, xy)$$

- Equality Conditions:

$$(x_1, y_1) = (x_2, y_2) \Rightarrow x_1 = x_2, y_1 = y_2$$

{Above are same rules as for 2D vectors.}

- Complex Conjugate:

$$\bar{z} = (x, -y) \qquad \{where \qquad z = (x, y)\}$$

- Multiplication:

$$(x_1, y_1)(x_2, y_2) = (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1)$$

{For vectors have "dot" and cross" products.}

- Absolute Value (Modulus):

$$|z| = \sqrt{z \cdot \bar{z}} = \sqrt{x^2 + y^2}$$

i.e., Pythagorean Theorem: length of "vector" z.

- Division

$$\frac{z_1}{z_2} = \frac{z_1 \cdot z_2}{z_2 \cdot z_2} = \frac{(x_1, y_1), (x_2, -y_2)}{x_2^2 + y_2^2}$$

$$= \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2}, \frac{x_2 y_1 - x_1 y}{x_2^2 + x_2^2}$$

Also:

- Any real number can be written in complex form as

$$a = (a, 0)$$

- “Pure imaginary number”: (0, b)

Solution to $x^2 + 1 = 0$
is then a complex number (a, b) satisfying:

$$(a, b)^2 + 1 = 0$$

$$(a^2 - b^2, 2ab) + (1, 0) = (0, 0)$$

with solution $a = 0, b = \pm 1$

or $x = \pm(0, 1)$.

Alternate Notation:

Denote $i = \sqrt{-1}$

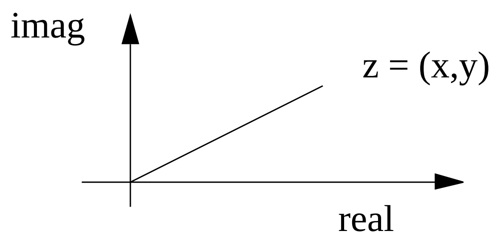
Then solution to $x^2 + 1 = 0$
is written $x = \pm i$.

And $i = (0, 1)$

with complex numbers now written as $z = x + iy$

{in EE, use $j = \sqrt{-1}$ }

Polar Coordinate Representation



r = |z|

$$z = r(\cos\theta + i\sin\theta)$$

or

$$z = re^{i\theta}$$ (Euler's Formula)

where $x = r\cos\theta$ $y = r\sin\theta$

Complex concepts can be extended to higher dimensions using

Quaternions:

$$z = x_0 + ix_1 + jx_2 + kx_3$$

$$i^2 = j^2 = k^2 = -1 \quad and \quad ij = -ji = k$$

For:

$x_3 = 0$ we have a 3D complex number,

$x_2 = x_3 = 0$ we have a 2D complex number.