Introduction to Numerical Methods

Numerical Analysis/Methods

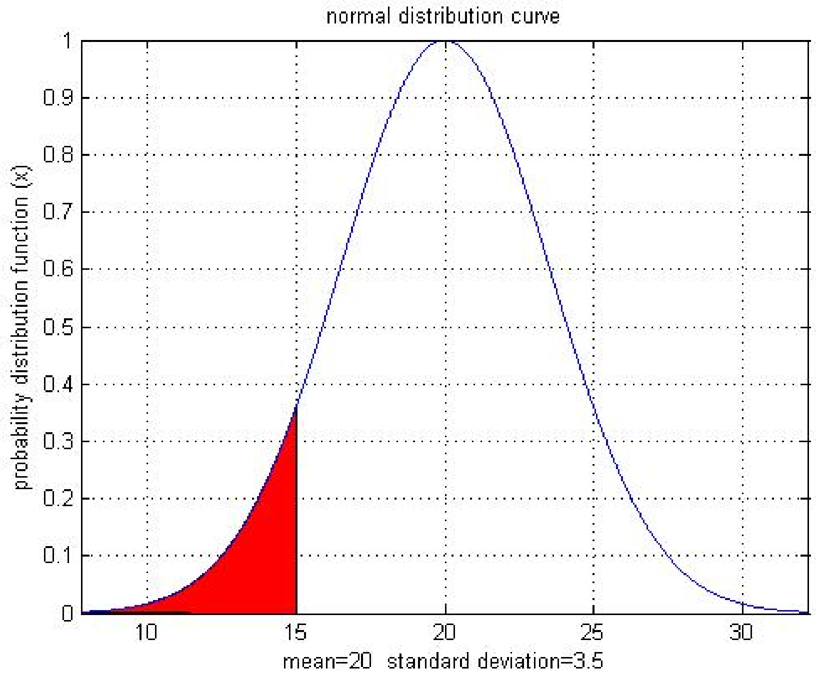
- What is numerical analysis/method?
  - Analysis and design of algorithms for numerically solving mathematical problems in science and engineering

- Why do we care about numerical analysis?
  - Simulation of real-world phenomena and events
  - Virtual prototyping of engineering designs

Why use Numerical Methods?

- To solve problems that cannot be solved exactly

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du$$



Solving an Engineering Problem

How do we solve an engineering problem?

Problem Description

↓

Mathematical Model

↓

Solution of Mathematical Model

↓

Using the Solution

Analysis vs. Numerical Analysis

- Consider solving $x^2=2$
- Analytically, we know that is a root of the equation
- Numerically, how do we find the root of the equation using a computer program?
- Computer can only do arithmetic operations
- Design a procedure consisting of only arithmetic operations to find the root

Mathematical Procedures

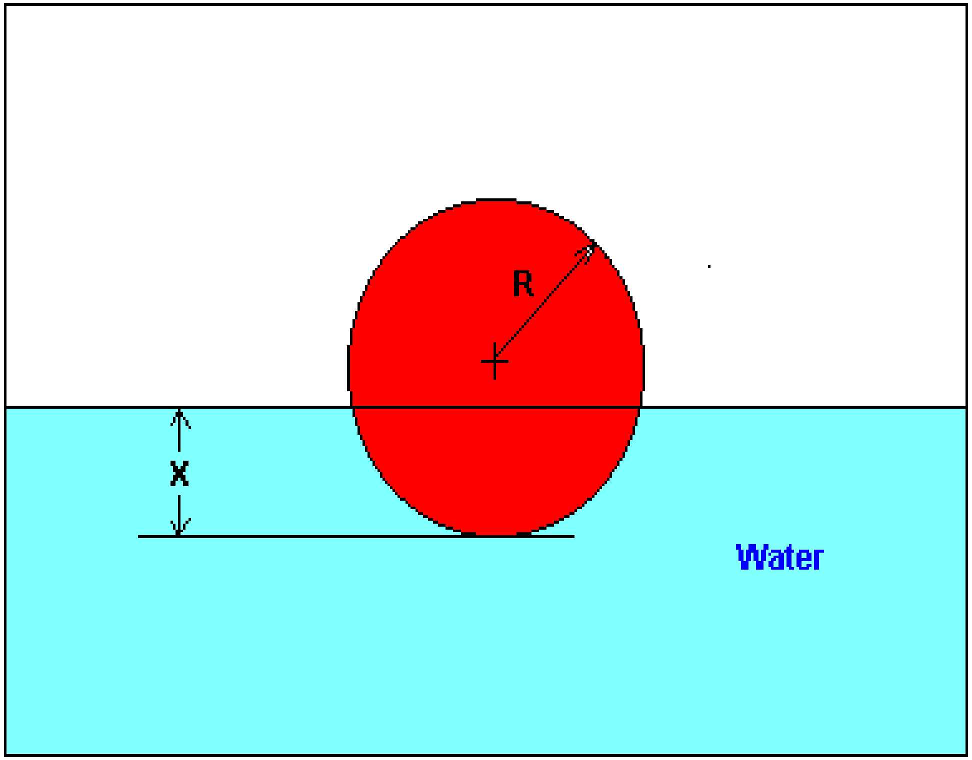
- Root finding (Nonlinear Equations)
  - Bracketing
  - Open
- Curve Fitting
  - Interpolation
  - Regression
- Differentiation/Integration
- Simultaneous Linear Equations
  - LU Decomposition
- Ordinary Differential Equations
- Partial Differential Equations

Nonlinear Equations

How much of the floating ball is under water?

Diameter=0.11m

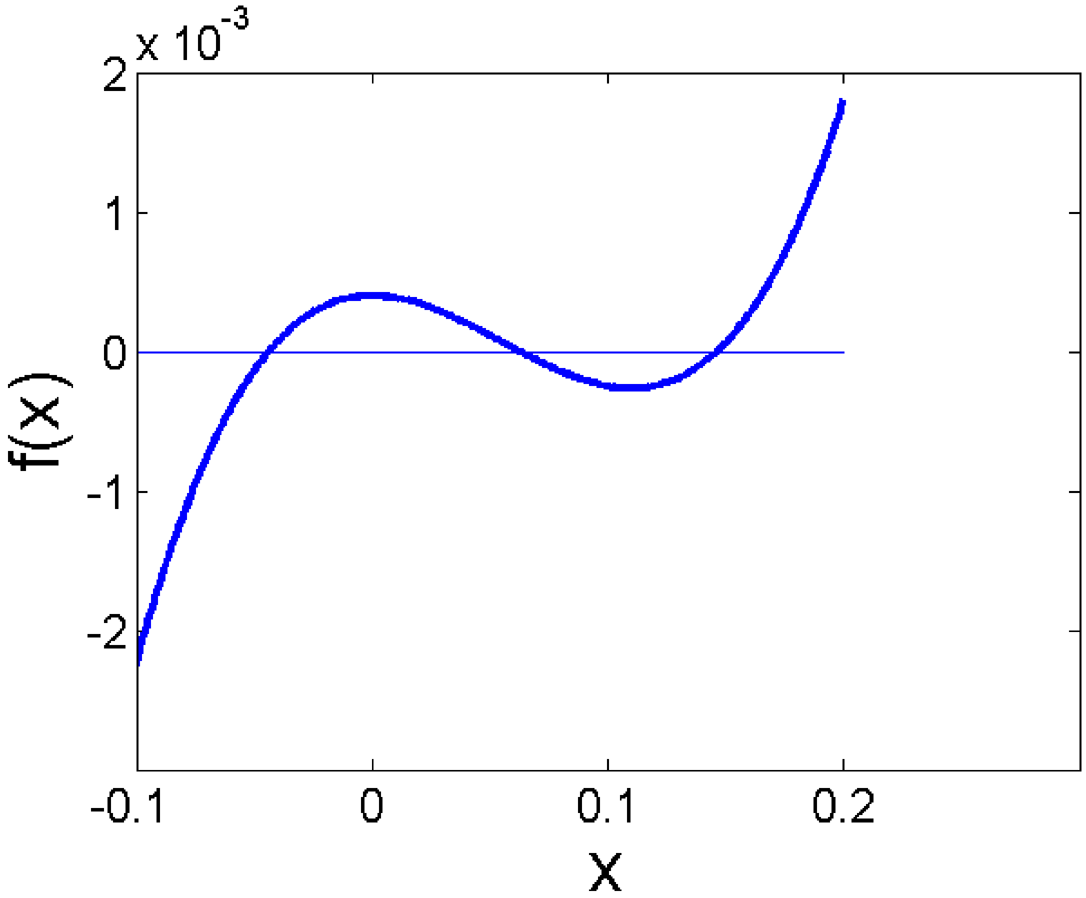
Specific Gravity=0.6



$$x^3 - 0.165x^2 + 3.993\times10^{-4} = 0$$

Nonlinear Equations

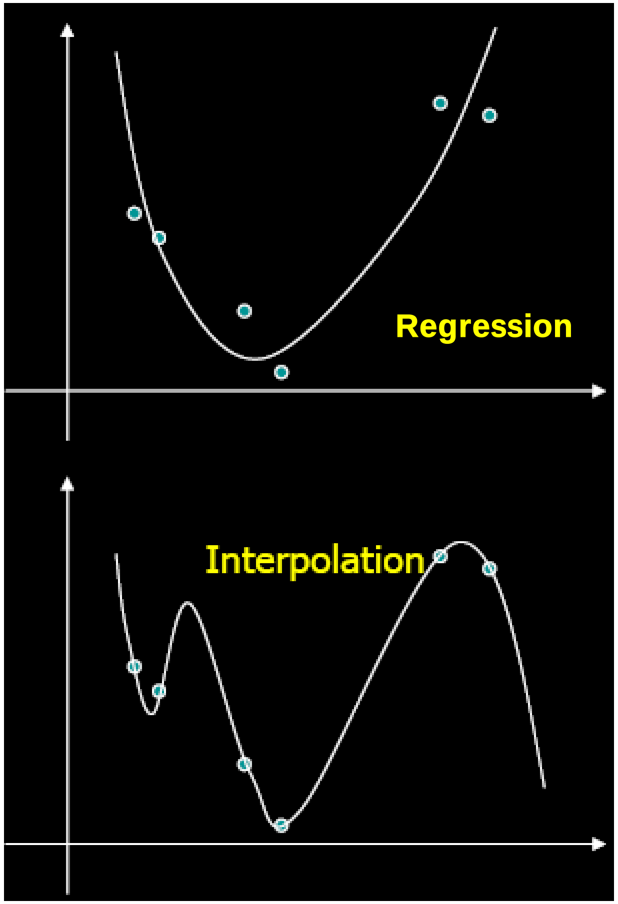
How much of the floating ball is under the water?



$$f(x) = x^3 - 0.165x^2 + 3.993\times10^{-4} = 0$$

Curve fitting

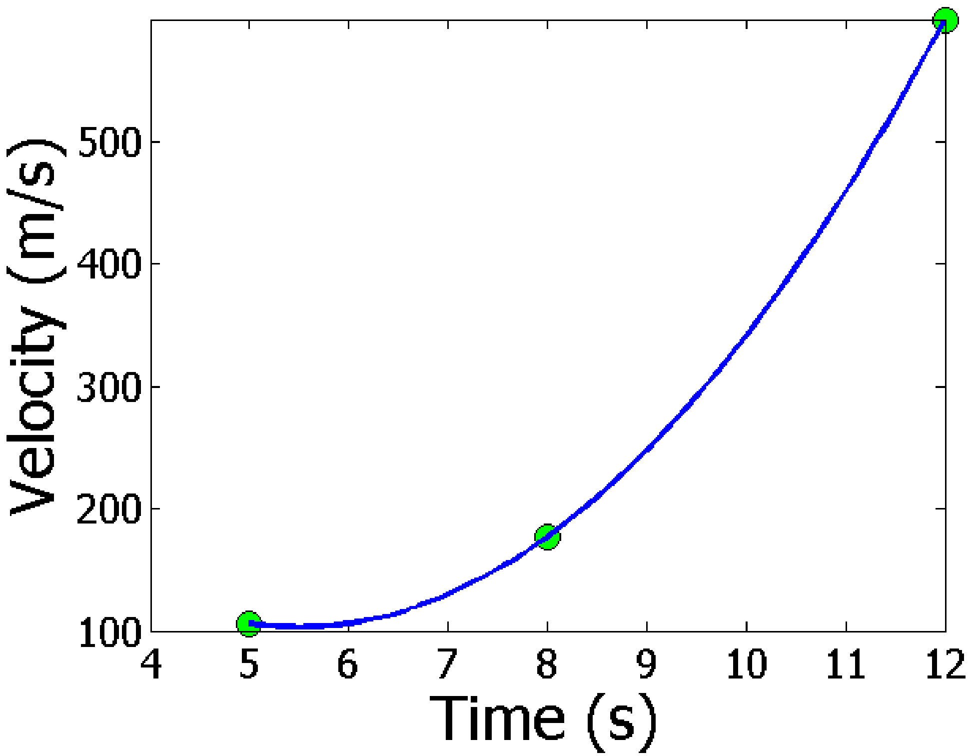
- Interpolation and curve fitting
- Find intermediate values from a table of data
- Fit curves to data
- If the curve passes all data points, we call it interpolation.



Interpolation

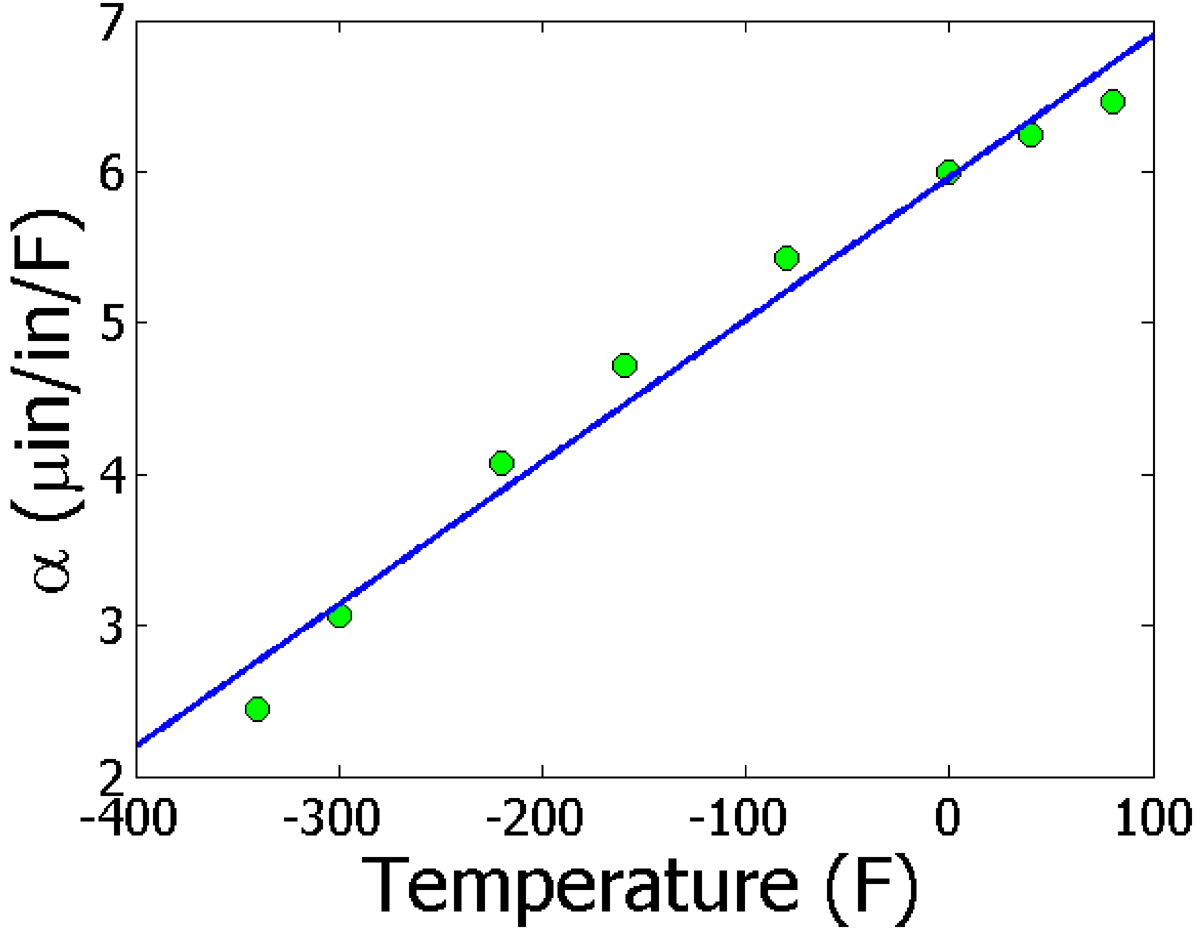
What is the velocity of the rocket at t=7 seconds?

Time (s)	5	8	12
Vel (m/s)	106	177	600



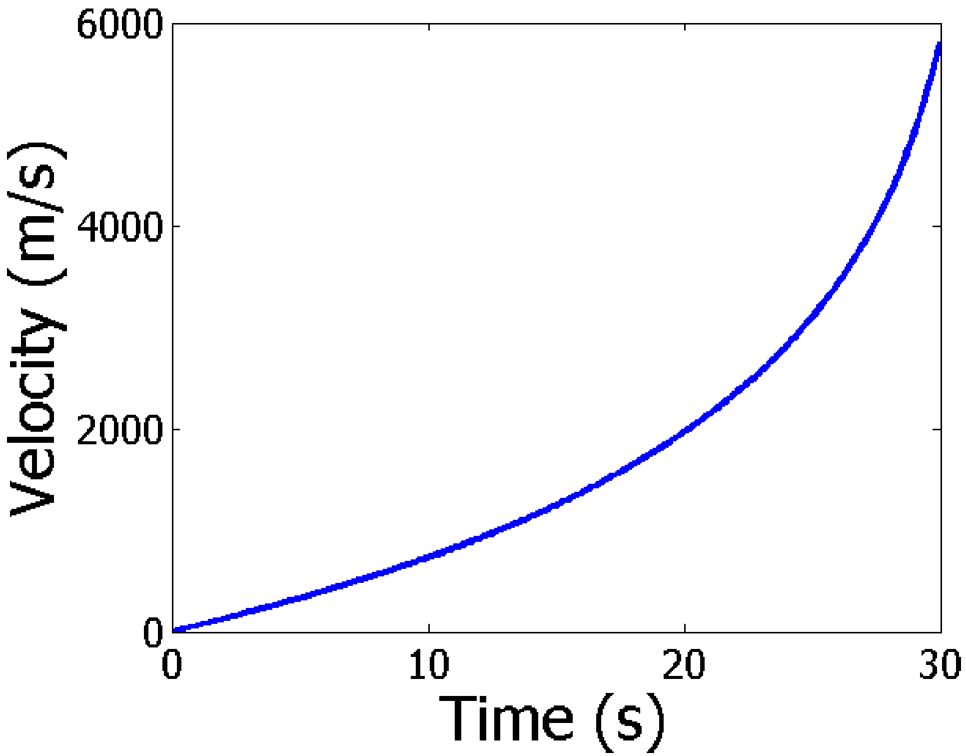
Regression

Thermal expansion coefficient data for cast steel



Differentiation

What is the acceleration at t=7 seconds?



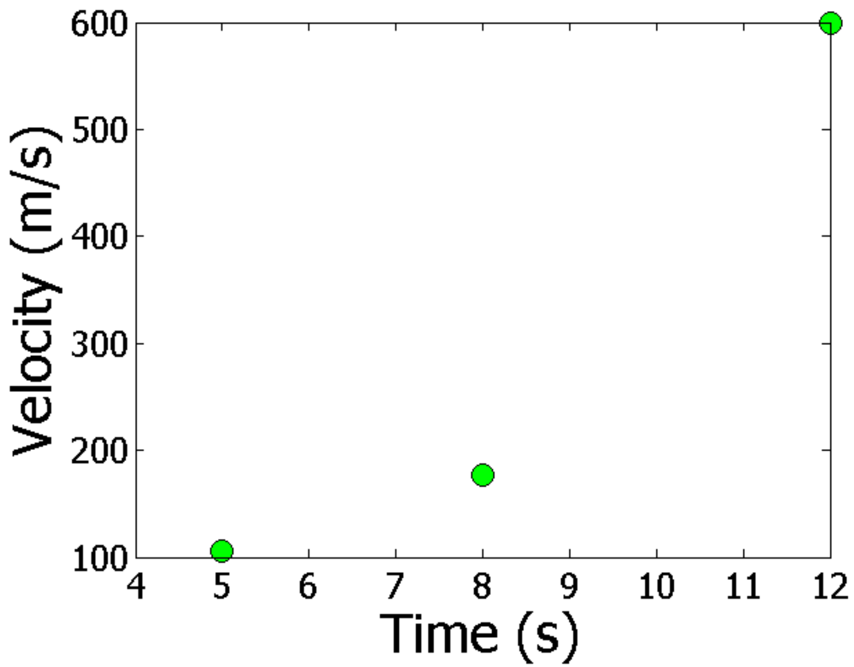
$$v(t) = 2200 \ \ln\left(\frac{16\times 10^4}{16\times 10^4 - 5000t}\right) - 9.8t$$

$$a = \frac{dv}{dt}$$

Differentiation

What is the acceleration at t=7 seconds?

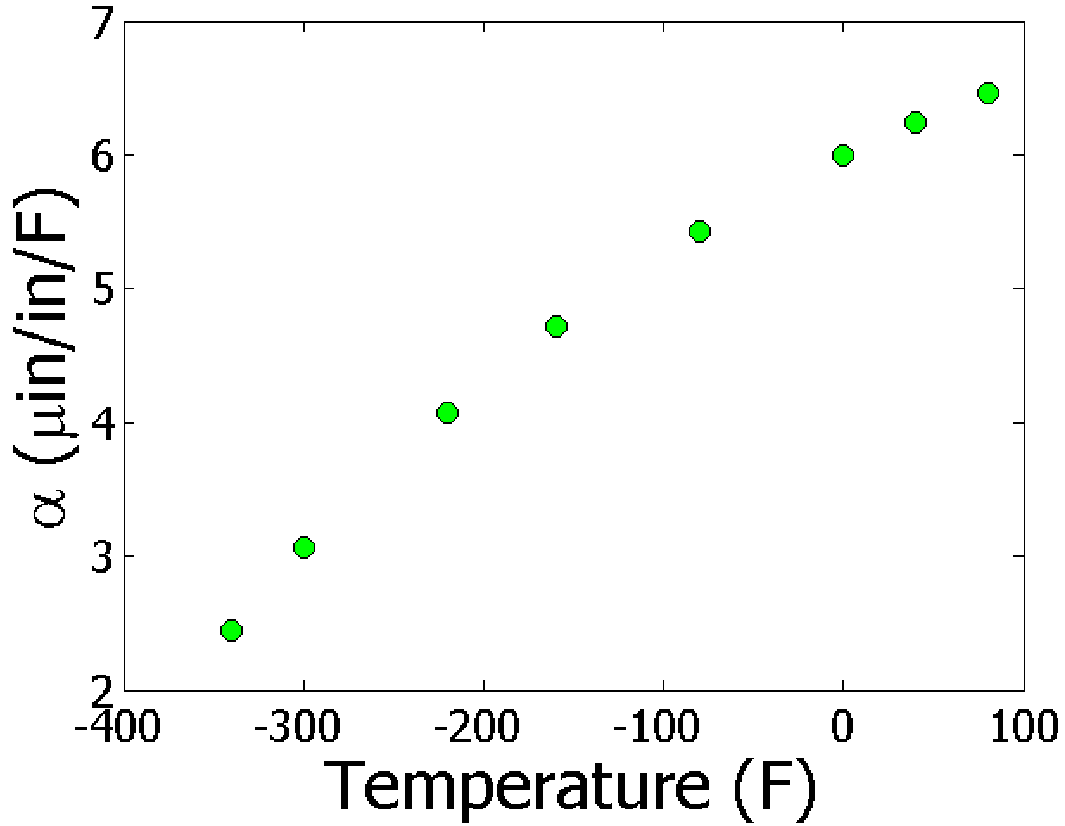
Time (s)	5	8	12
Vel (m/s)	106	177	600



Integration

Finding the diametric contraction in a steel shaft when dipped in liquid nitrogen.

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



Simultaneous Linear Equations

Find the velocity profile, given

Time (s)	5	8	12
Vel (m/s)	106	177	600

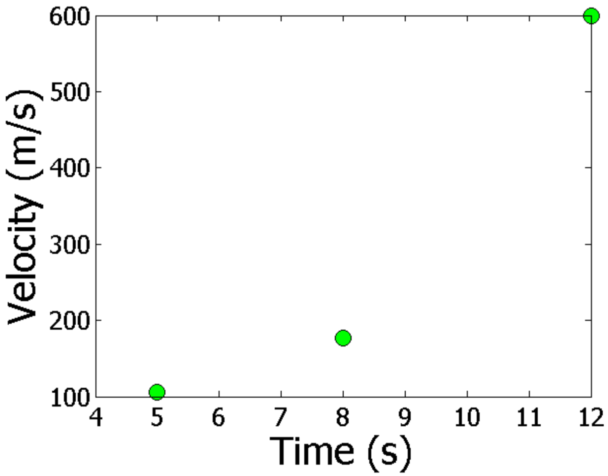
$$v(t) = at^2 + bt + c, \; 5 \le t \le 12$$

Three simultaneous linear equations

$$25a + 5b + c = 106$$

$$64a + 8b + c = 177$$

$$144a + 12b + c = 600$$



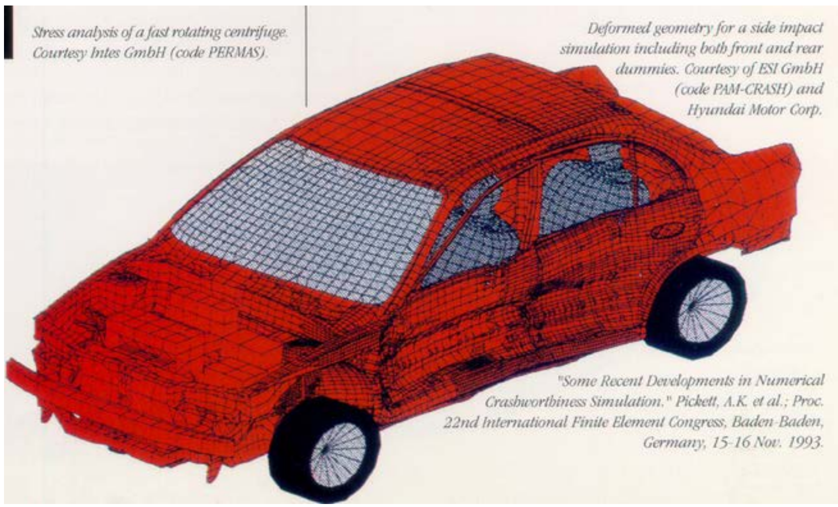
Ordinary Differential Equations

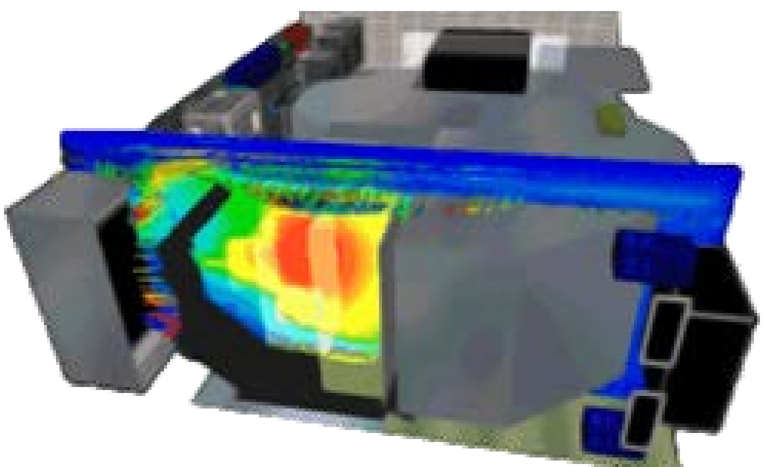
How long does it take a trunnion to cool down?

$$mc\frac{d\theta}{dt} = -hA(\theta - \theta_a), \quad \theta(0) = \theta_{room}$$

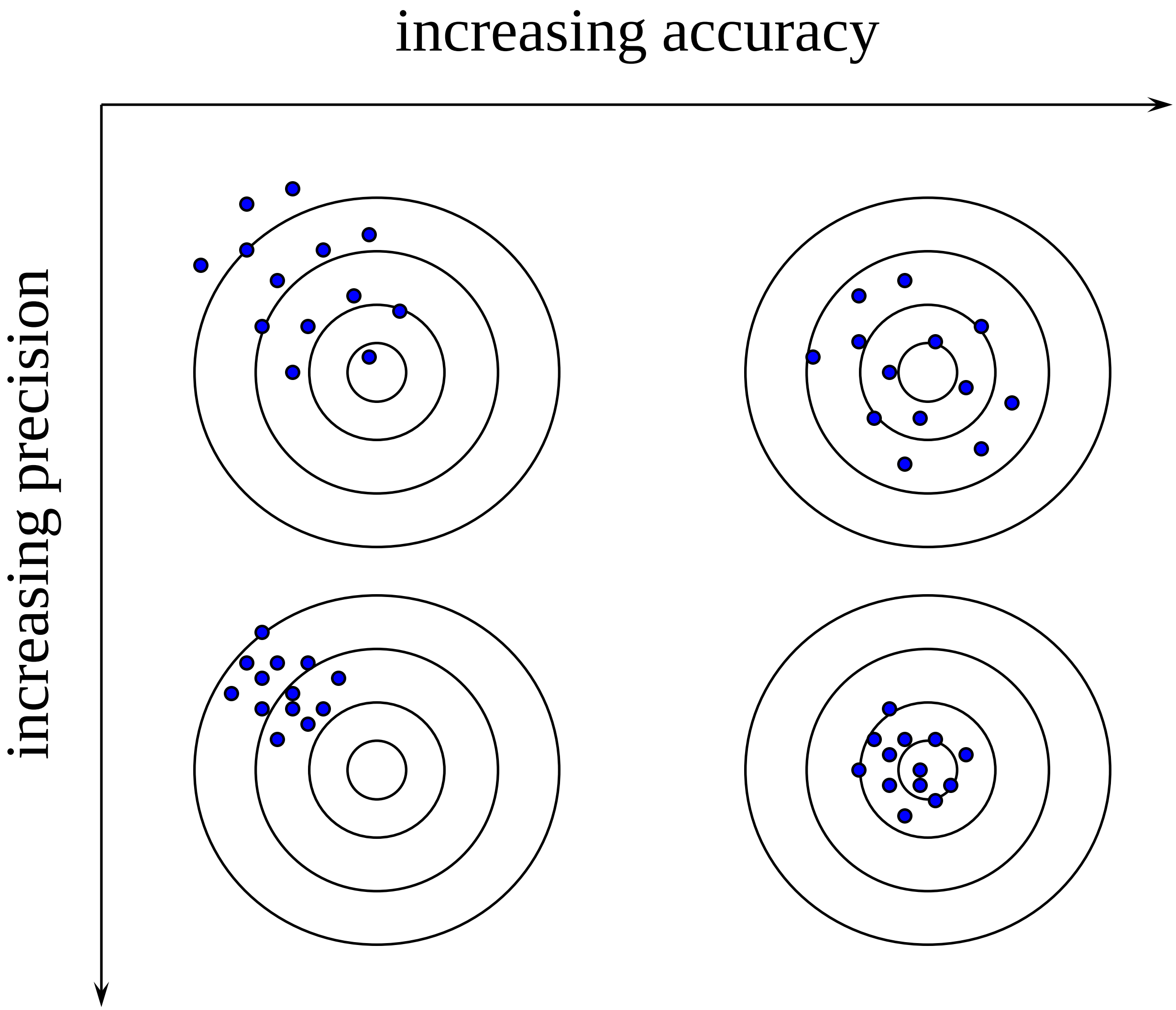
Partial Differential Equations

PDEs are used to model many systems in many different fields of science and engineering. Modelling in Industry: Automobiles, Aerospace, Electronics





Accuracy and Precision



Error Definitions

- Round off error - originate from the fact that computers retain only a fixed number of significant figures caused by representing a number approximately

$$\frac{1}{3} \cong 0.333333 \qquad \sqrt{2} \cong 1.4142...$$

- Truncation errors - errors that result from using an approximation in place of an exact mathematical procedure. Taking only a few terms of e^x

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

If only 3 terms are used,

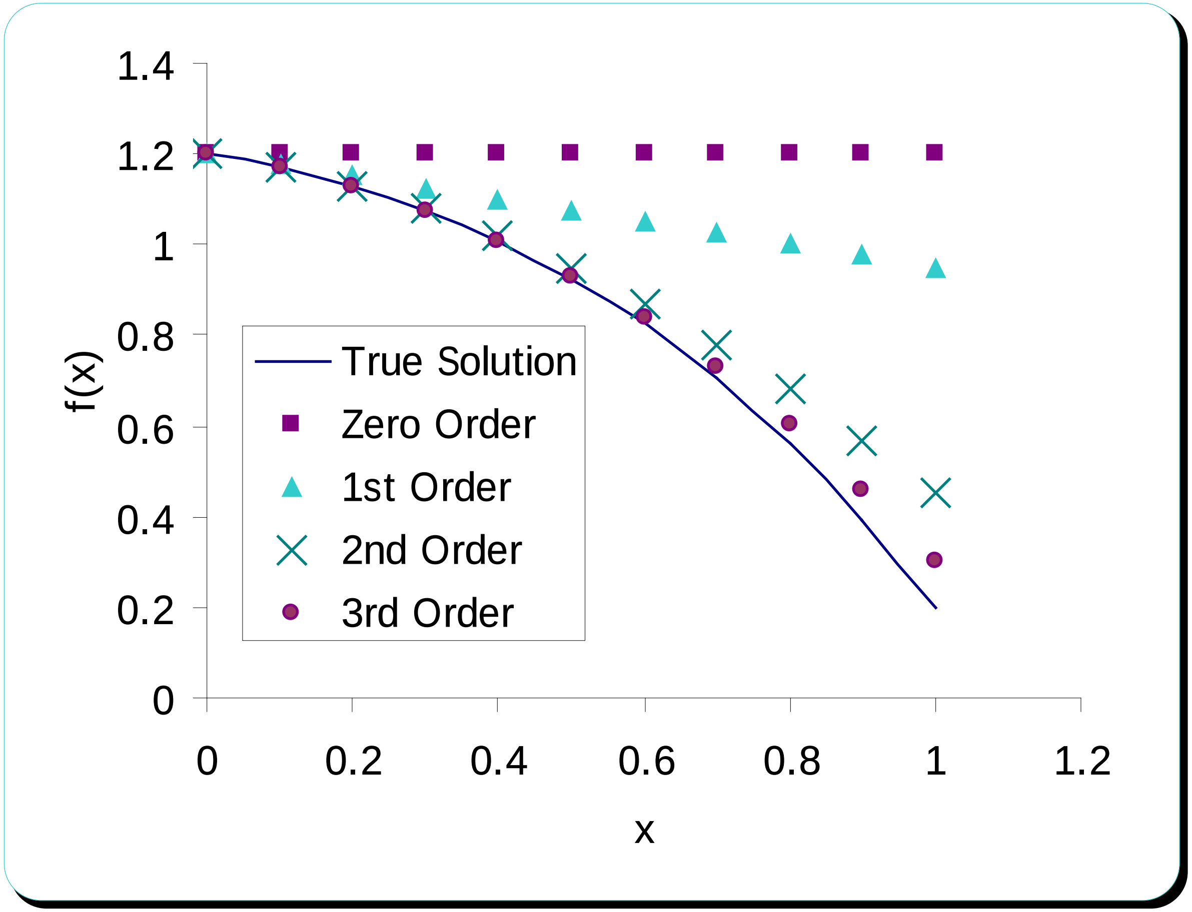
$$Truncation \;\; Error = e^x - \left(1 + x + \frac{x^2}{2!}\right)$$

Taylor Series Expansion

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(x_i)}{3!}h^3 + \ldots\ldots + \frac{f^{n}(x_i)}{n!}h^n + R_n$$

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1} \quad x_i < \xi < x_{i+1} \quad where \;\; h = step \;\; size = x_{i+1} - x_i$$

Solution



True Error

- Defined as the difference between the true value in a calculation and the approximate value found using a numerical method etc.

True Error = True Value – Approximate Value

Relative True Error

- Defined as the ratio between the true error, and the true value.

$$\text{Relative True Error } (\in_t) = \frac{\text{True Error}}{\text{True Value}}$$

Approximate Error

- Approximate error is defined as the difference between the present approximation and the previous approximation.

Approximate Error (E_a) = Present Approximation – Previous Approximation

Relative Approximate Error

- Defined as the ratio between the approximate error and the present approximation.

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

How a Decimal Number is Represented?

$$257.76 = 2\times10^{2} + 5\times10^{1} + 7\times10^{0} + 7\times10^{-1} + 6\times10^{-2}$$

Base 2

$$(1011.0011)_2 = \begin{pmatrix} (1\times2^{3} + 0\times2^{2} + 1\times2^{1} + 1\times2^{0}) \\ +(0\times2^{-1} + 0\times2^{-2} + 1\times2^{-3} + 1\times2^{-4}) \end{pmatrix}_{10} = 11.1875$$

Converting a base-10 integer to binary representation.

Number	Quotient	Remainder
11/2	5	$1 = a_0$
5/2	2	$1 = a_1$
2/2	1	$0 = a_2$
1/2	0	$1 = a_3$

Hence $$(11)_{10} = (a_3 a_2 a_1 a_0)_2 = (1011)_2$$

Converting a base-10 fraction to binary representation.

Fractional	Number	After decimal	Before decimal
0.1875×2	0.375	0.375	$0 = a_{-1}$
0.375×2	0.75	0.75	$0 = a_{-2}$
0.75×2	1.5	0.5	$1 = a_{-3}$
0.5×2	1.0	0.0	$1 = a_{-4}$

Hence $(0.1875)_{10} = (a_{-1}a_{-2}a_{-3}a_{-4})_2 = (0.0011)_2$

Converting a base-10 fraction to approximate binary representation.

	Number	**Number after decimal**	**Number before Decimal**
0.3×2	0.6	0.6	$0 = a_{-1}$
0.6×2	1.2	0.2	$1 = a_{-2}$
0.2×2	0.4	0.4	$0 = a_{-3}$
0.4×2	0.8	0.8	$0 = a_{-4}$
0.8×2	1.6	0.6	$1 = a_{-5}$

$$(0.3)_{10} \approx (a_{-1}a_{-2}a_{-3}a_{-4}a_{-5})_2 = (0.01001)_2 = 0.28125$$

Another Way to Look at Conversion

Convert $(11.1875)_{10}$ to base 2

$$(11)_{10} = 2^3 + 3 = 2^3 + 2^1 + 1 = 2^3 + 2^1 + 2^0$$

$$= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

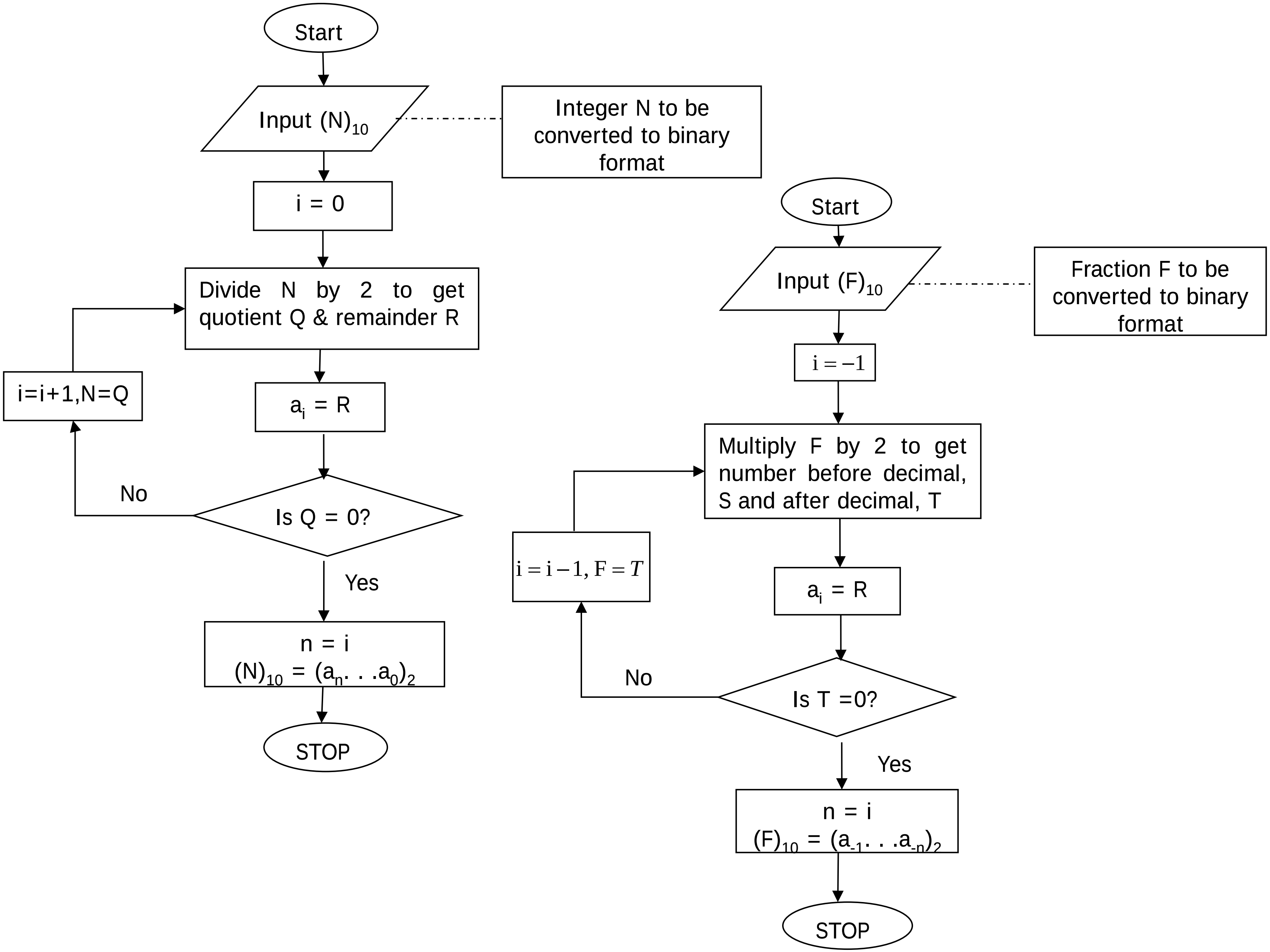
$$= (1011)_2$$

$$(0.1875)_{10} = 2^{-3} + 0.0625 = 2^{-3} + 2^{-4}$$

$$= 0 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} + 1 \times 2^{-4}$$

$$= (.0011)_2$$

$$(11.1875)_{10} = (1011.0011)_2$$


Start
Input $(N)_{10}$
Integer N to be converted to binary format
i = 0
Divide N by 2 to get quotient Q & remainder R
a_i = R
Is Q = 0?
No
i=i+1,N=Q
Yes
n = i
$(N)_{10} = (a_n...a_0)_2$
STOP
Start
Input $(F)_{10}$
Fraction F to be converted to binary format
$i = -1$
Multiply F by 2 to get number before decimal, S and after decimal, T
a_i = R
Is T =0?
No
$i = i - 1, F = T$
Yes
n = i
$(F)_{10} = (a_{-1}...a_{-n})_2$
STOP

Floating Decimal Point
Scientific Form

$$256.78 \text{ is written as } +2.5678\times 10^{2}$$

$$0.003678 \text{ is written as } +3.678\times 10^{-3}$$

$$-256.78 \text{ is written as } -2.5678\times 10^{2}$$

Example : The form is $\text{sign}\times\text{mantissa}\times 10^{\text{exponent}}$

or $\sigma\times m\times 10^{e}$

Example: For -2.5678×10^{2}

$$\sigma=-1$$

$$m=2.5678$$

$$e=2$$

Floating Point Format for Binary Numbers

$$y = \sigma \times m \times 2^{e}$$

$$\sigma = \text{sign of number } \left(0 \text{ for } + \text{ve}, 1 \text{ for - ve}\right)$$

$$m = \text{mantissa} \left[(1)_2 < m < (10)_2\right]$$

1 is not stored as it is always given to be 1.

$$e = \text{integer exponent}$$

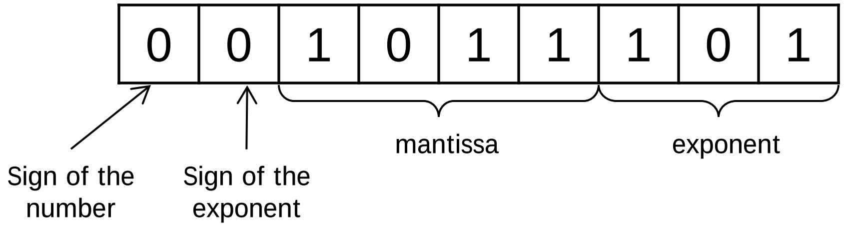
Example

9 bit-hypothetical word

- the first bit is used for the sign of the number,
- the second bit for the sign of the exponent,
- the next four bits for the mantissa, and
- the next three bits for the exponent

$$(54.75)_{10} = (110110.11)_2 = (1.1011011)_2 \times 2^5$$
$$\cong (1.1011)_2 \times (101)_2$$

We have the representation as



IEEE-754 Format Single Precision

32 bits for single precision $\text{Value} = (-1)^{s} \times (1\bullet m)_2 \times 2^{e'-127}$

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Sign(s) Biased Exponent (e’) Mantissa (m)

Example

1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$$\begin{aligned}
\text{Value} &= (-1)^{s} \times (1.m)_2 \times 2^{e'-127} \\
&= (-1)^{1} \times (1.10100000)_2 \times 2^{(10100010)_2-127} \\
&= (-1) \times (1.625) \times 2^{162-127} \\
&= (-1) \times (1.625) \times 2^{35} = -5.5834 \times 10^{10}
\end{aligned}$$