Mathematical Preliminaries

Vasily Arzhanov

Reactor Physics, KTH

Approximation Error

True value: a

Approximate value: $\tilde{a}$

$$a \approx \tilde{a}$$

Approximation Error $\Delta a \equiv \tilde{a} - a$

Absolute Error is $\left|\Delta a\right|$

An upper bound is any (known) number Δ_a such that $\left|\Delta a\right| \leq \Delta_a$

$$\tilde{a} - \Delta_a \leq a \leq \tilde{a} + \Delta_a \longrightarrow a = \tilde{a} \pm \Delta_a$$

Acceleration in Sweden

Acceleration g in Sweden $\quad 9.81666 \le g \le 9.82008$

Best guess $\quad \tilde{g} = \dfrac{9.82008 + 9.81666}{2} = 9.81837$

Uncertainty $\quad \Delta_g = \dfrac{9.82008 - 9.81666}{2} = 0.00171$

Neutron Mass

NIST reports $\tilde{m} = 1.674\ 927\ 498\ 04 \times 10^{-27}$ kg

$\Delta_m = 0.000\ 000\ 000\ 95 \times 10^{-27}$ kg

$\tilde{m} = 1.674\ 927\ 498\ 04(95) \times 10^{-27}$ kg

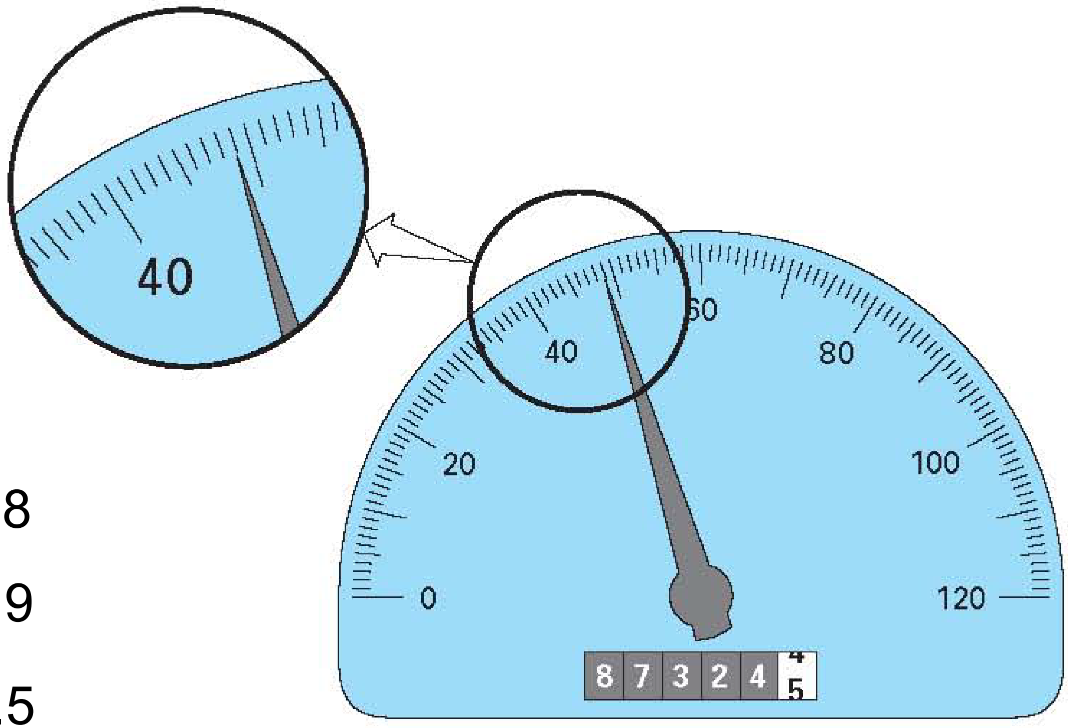
(Exact uncertainty) $|\Delta m| \le \Delta_m = 0.000\ 000\ 000\ 95 \times 10^{-27}\,\text{kg}$

$$\tilde{m} - \Delta_m \le m \le \tilde{m} + \Delta_m$$

Sensor Readings

Visual inspection:

$$48 \leq v \leq 49$$



One person insists: v = 48.8

Another insists: v = 48.9

Commonly accepted: v = 48.5

The estimated digit is one-half of the smallest scale division.

Significant Digits

The significant digits of a number are those that can be used with confidence.

Significant digits are certain digits plus one estimated digit.

$v = 48.5$

Zeros are not always significant: 0.2021, 0.02021, 0.002021

Unclear: 202100 may have 4, 5 or 6 significant digits.

Resolve: 2.02100×10^5 (6 significant digits).

Round-off Errors

Computers may retain only limited numbers of digits.

Specific numbers: $\sqrt{2},\ \pi, e, \ldots$ have infinitely many significant digits.

$$\pi = 3.14159265358979323846264 3\ldots$$

The omission of the remaining significant figures is called round-off error.

Two Rounding Rules

Chopping: $1.650 \approx 1.6$

Nearest: $1.650 \approx 1.7$

x	Chop	Nearest	$x-\tilde{x}$ Chop	$x-\tilde{x}$ Nearest
1.649	1.6	1.6	0.049	0.049
1.650	1.6	1.7	0.050	-0.050
1.651	1.6	1.7	0.051	-0.049
1.699	1.6	1.7	0.099	-0.001
1.749	1.7	1.7	0.049	0.049
1.750	1.7	1.8	0.050	-0.050

Relative Error

$$l = \quad 1\,\text{cm} \pm 1\,\text{cm}$$
$$l = 100\,\text{cm} \pm 1\,\text{cm}$$

$$\forall a \neq 0 \quad \delta \equiv \frac{\Delta a}{a} = \frac{\tilde{a} - a}{a} \longrightarrow \tilde{a} = a(1+\delta)$$

An upper bound is any (known) number δ_a such that $|\delta| \leq \delta_a$

$$|\delta| = \frac{|\Delta a|}{|a|} \longrightarrow |\Delta a| = |a| \cdot |\delta| \leq |a|\,\delta_a \longrightarrow \Delta_a = |a|\,\delta_a$$

$$\Delta_a = |a|\,\delta_a \approx |\tilde{a}|\,\delta_a \longrightarrow \tilde{a}(1-\delta_a) \leq a \leq \tilde{a}(1+\delta_a) \longrightarrow a = \tilde{a}(1 \pm \delta_a)$$

Approximation Error of a Sum

$$\tilde{x}_i = x_i + \Delta x_i \qquad \left|\Delta x_i\right| \leq \Delta_i$$

$$\tilde{x} = \tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n = x + \Delta x$$

$$\Delta x = \Delta x_1 + \Delta x_2 + \ldots + \Delta x_n$$

$$\left|\Delta x\right| \leq \Delta_1 + \Delta_2 + \ldots + \Delta_n = \Delta_x$$

Relative Error of a Sum

$$\text{All } x_i > 0; \qquad |\Delta x_i| \le \Delta_i; \qquad \frac{|\Delta x_i|}{x_i} \le \delta_i; \qquad \delta_{\max} \equiv \max \delta_i$$

$$\delta \equiv \frac{\Delta x_1 + \Delta x_2 + \ldots + \Delta x_n}{x_1 + x_2 + \ldots + x_n}$$

$$|\delta| \le \frac{x_1\delta_1 + x_2\delta_2 + \ldots + x_n\delta_n}{x_1 + x_2 + \ldots + x_n} \le \delta_{\max}$$

Relative Error of Product

$$\tilde{x} = \tilde{x}_1 \cdot \tilde{x}_2 \cdot \ldots \cdot \tilde{x}_n; \qquad \text{All } x_i > 0;$$

$$\delta \equiv \left| \frac{\Delta x}{x} \right| \leq \delta_1 + \delta_2 + \ldots + \delta_n$$

Selected Cases

$$\tilde{u} = k \cdot \tilde{x} \longrightarrow \delta_u = \delta_x; \quad \Delta u = k \cdot \Delta x$$

$$\tilde{u} = \tilde{x}_1 / \tilde{x}_2 \longrightarrow \delta_u = \delta_1 + \delta_2$$

$$\tilde{u} = \tilde{x}^m \longrightarrow \delta_u = m\delta_x$$

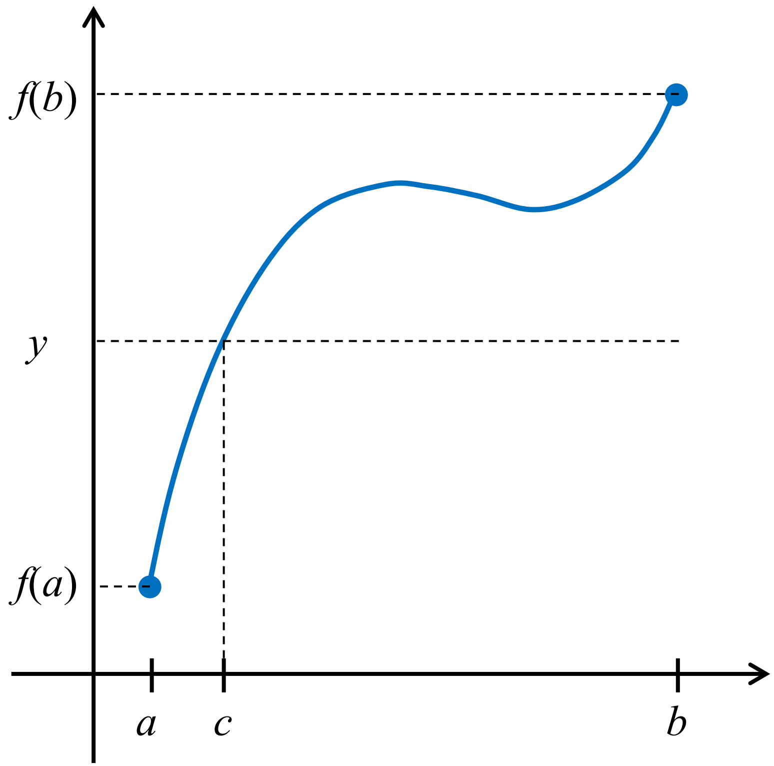
$$\tilde{u} = \sqrt[m]{\tilde{x}} \longrightarrow \delta_u = \frac{1}{m}\delta_x$$

Intermediate Value Theorem

Continuous function

$$\lim_{x \to c} f(x) = f(c)$$

Let $f(x)$ be a continuos function on $[a,b]$ then f realises every value between $f(a)$ and $f(b)$. More precisely, if y is a number between a and b, then there exists a number c, $a \le c \le b$, such that $y = f(c)$.



Continuous Limit Theorem

Let $f(x)$ be a continuos function in a neighborhood of x_0 and $\lim_{n\to\infty} x_n = x_0$ then

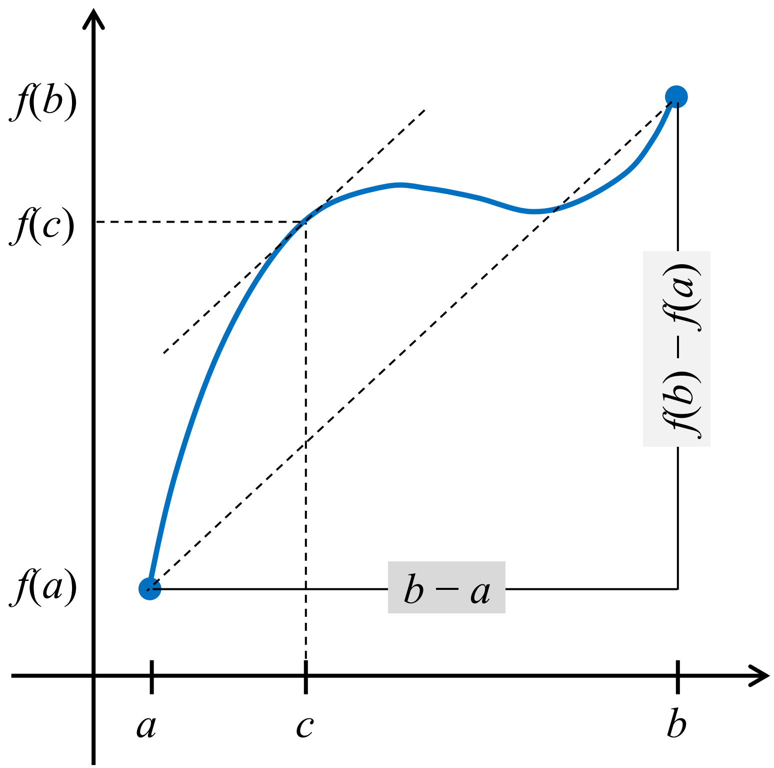
$$\lim_{n\to\infty} f\left(x_n\right) = f\left(\lim_{n\to\infty} x_n\right) = f\left(x_0\right)$$

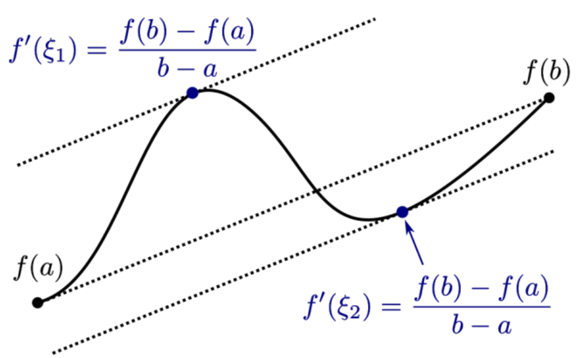
More precisely, limits may be brought inside continous functions.

Mean Value Theorem

Let $f(x)$ be a continuosly differentiable function on $[a, b]$. Then there exists a number c between a and b such that

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

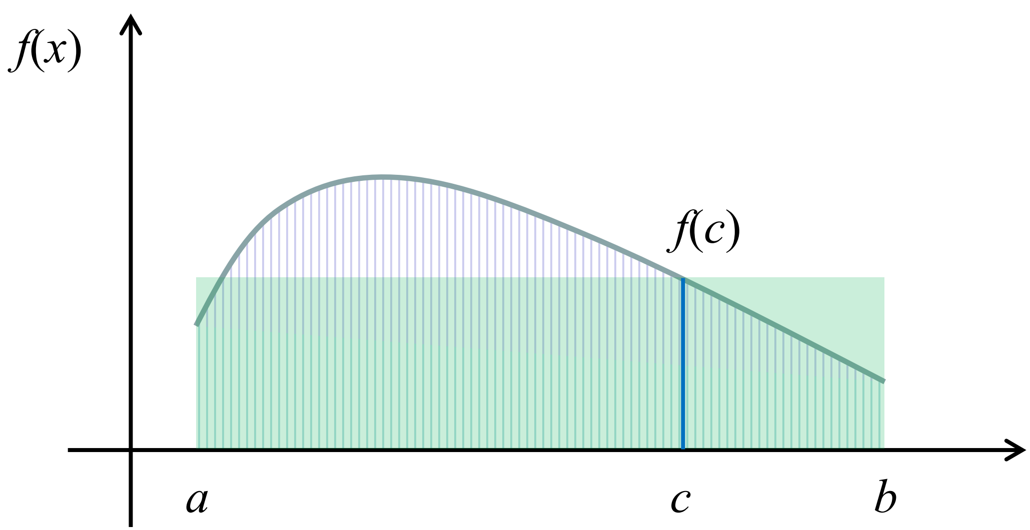




There could be several such numbers.

Mean Value for Integrals

Let $f(x)$ be a continuos function on a closed bounded interval $[a,b]$ then there exists at least one number c such that $\int_a^b f(x)dx = f(c)(b-a)$



Mean-Value Theorem

$$f(x), g(x) \in C\left[a,b\right] \qquad g(x) \geq 0 \quad \forall x$$

$$\int_a^b f(x)g(x)dx = f(c)\int_a^b g(x)dx$$

$$\int_a^b f(x)dx = f(c)\left(b-a\right)$$

Uncertainty Sources

Input data

`x = 0.1`	$\text{x} := \tilde{x} = 0.1 + \Delta(0.1)$	Representation
`x = 9.81`	$\text{x} := \tilde{x} = g_{\text{true}} + \Delta(g)$	Experimental
`x = sqrt(a)`	$\text{x} := \tilde{x} = \sqrt{a} + \Delta(\sqrt{a})$	Calculations

`y = F(x)`

$$y = F(x)$$

Error propagation $y + \Delta y = F(x + \Delta x)$

How uncertainty in y is related to uncertainty in x ?

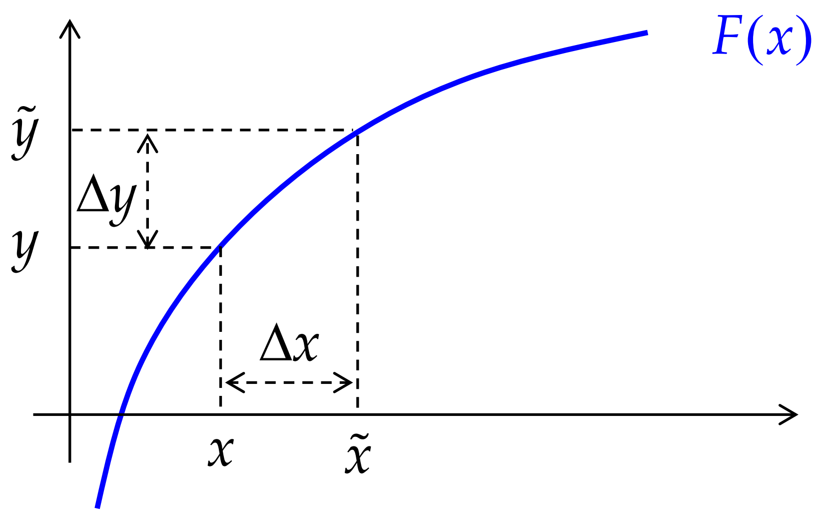
Assumption: no rounding errors (exact arithmetic)

Error Bounds

$$\tilde{x} = x + \Delta x \qquad |\Delta x| \le \Delta_x$$ Error bound in input

$$\tilde{y} = y + \Delta y \qquad |\Delta y| \le \Delta_y$$ Error bound in output

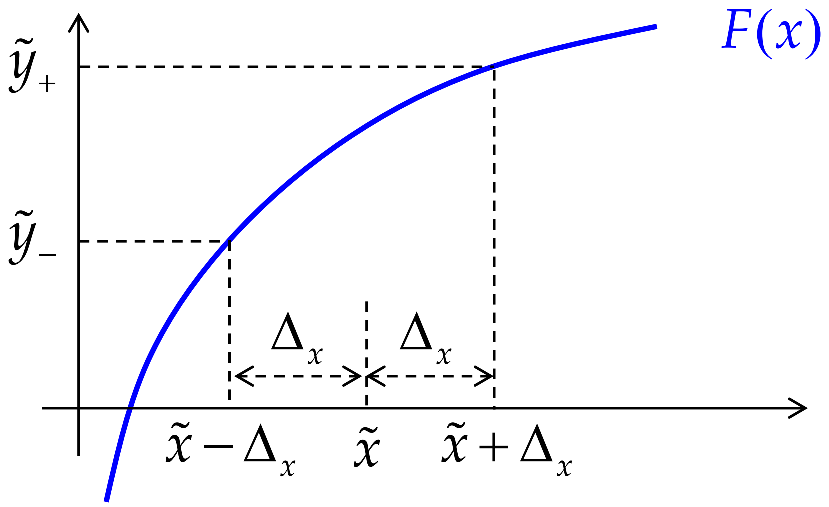
Notation $x = \tilde{x} \pm \Delta_x \qquad y = \tilde{y} \pm \Delta_y$



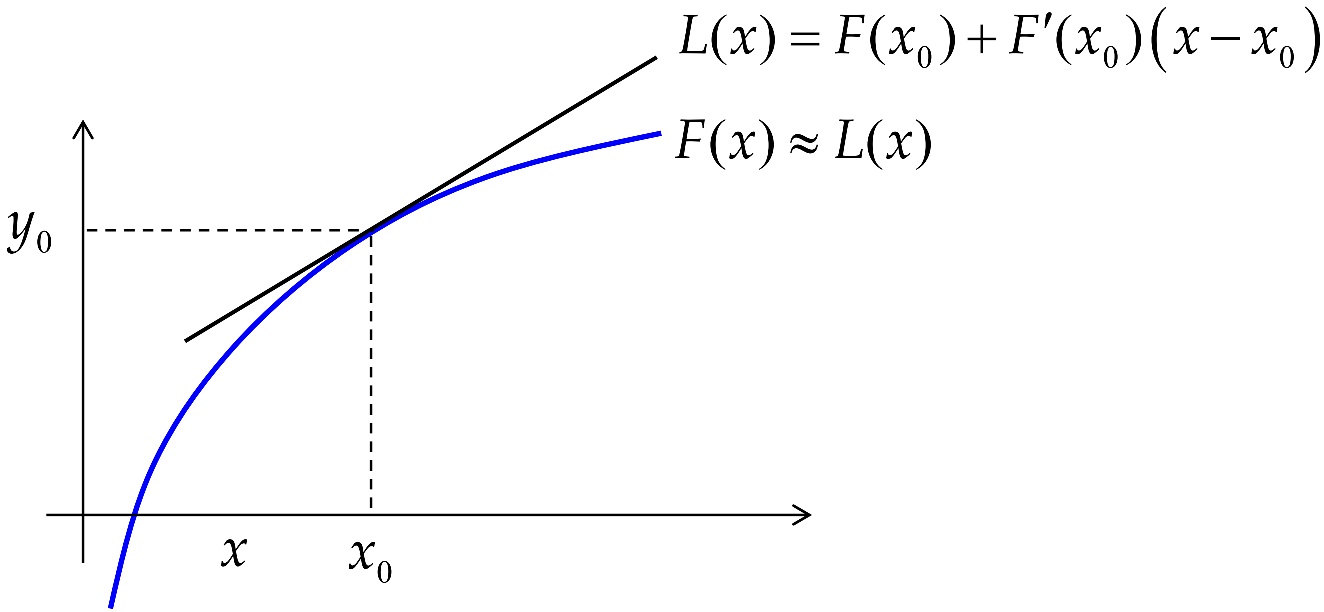
Exact Error Estimate

$$\tilde{x} = x + \Delta x \qquad \left|\Delta x\right| \le \Delta_x$$

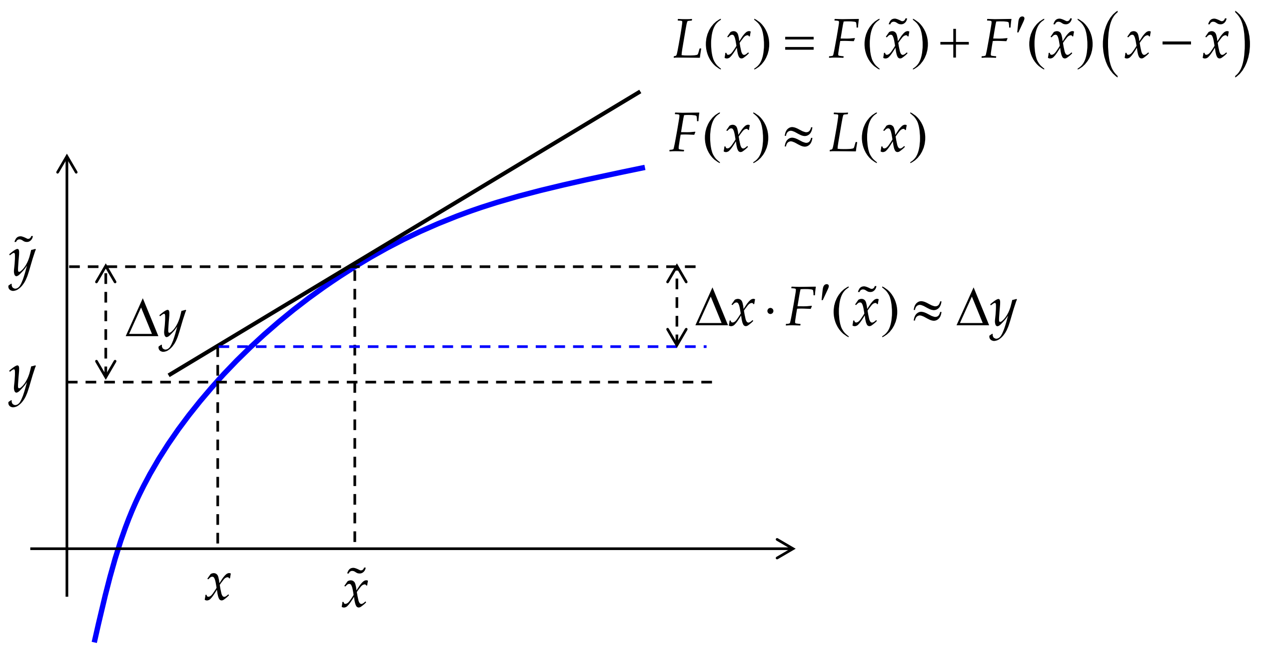
$$\tilde{y} = y + \Delta y \qquad \left|\Delta y\right| \le \Delta_y$$



Linearization



Approximate Error Estimate



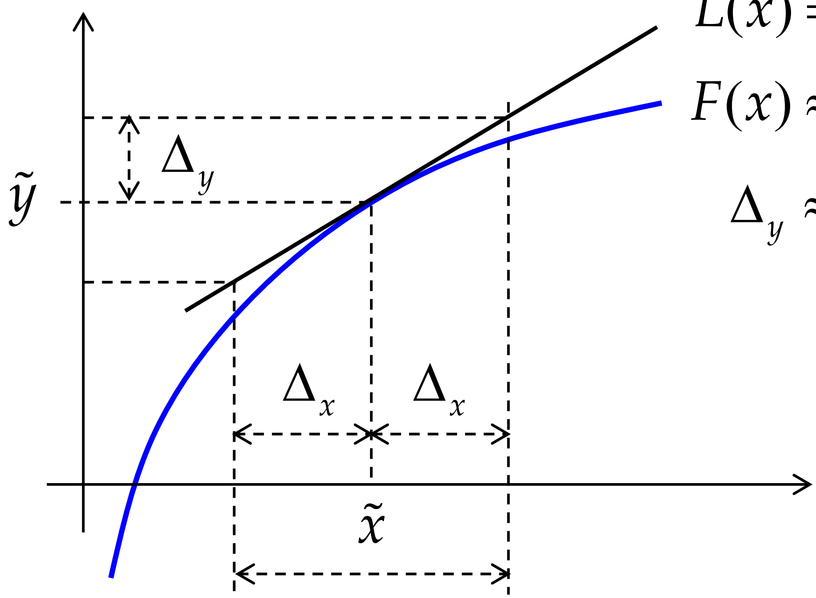
Error propagation $\Delta y \approx \Delta x \cdot F'(\tilde{x})$

Approximate Bounds

$$|\Delta x| \leq \Delta_x$$

$$\Delta x \in [-\Delta_x, \Delta_x]$$

$$x \in [\tilde{x} - \Delta_x, \tilde{x} + \Delta_x]$$



$$L(x) = F(\tilde{x}) + F'(\tilde{x})(x - \tilde{x})$$

$$F(x) \approx L(x)$$

$$\Delta_y \approx \Delta_x \cdot |F'(\tilde{x})|$$

Error propagation $\quad \Delta_y \approx \Delta_x \cdot |F'(\tilde{x})|$

Conditioning

- Well-conditioned: small errors in x induce small errors in y
  Condition: $|F'(\tilde{x})|$ is moderate.
- Ill-conditioned: small errors in x induce large errors in y
  Condition: $|F'(\tilde{x})|$ is large.

$$\Delta_y \approx \Delta_x \cdot |F'(\tilde{x})|$$

Relative error in x: $\delta_x \equiv \Delta_x / |x| \approx \Delta_x / |\tilde{x}|$

$$\delta_y \approx \frac{\Delta_y}{|\tilde{y}|} \approx \frac{|F'(\tilde{x})| \cdot \Delta_x}{|\tilde{y}|} = |F'(\tilde{x})| \frac{|\tilde{x}|}{|\tilde{y}|} \frac{\Delta_x}{|\tilde{x}|} = \kappa \delta_x$$

Relative error in y: $\delta_y \equiv \Delta_y / |y| \approx \Delta_y / |\tilde{y}|$

$$\kappa \equiv |F'(\tilde{x})| \frac{|\tilde{x}|}{|\tilde{y}|} \longrightarrow \delta_y \approx \kappa \delta_x$$

Multivariate Functions

Bivariate $\quad z = F(x,y) \quad \Delta z \approx \Delta x \cdot F_x(\tilde{x},\tilde{y}) + \Delta y \cdot F_y(\tilde{x},\tilde{y})$

$$\Delta_z \approx \Delta_x \cdot \left|F_x(\tilde{x},\tilde{y})\right| + \Delta_y \cdot \left|F_y(\tilde{x},\tilde{y})\right|$$

$$|\Delta z| \le \Delta_z$$

Multivariate $\quad z = F(x_1, x_2, \ldots, x_n)$

$$\Delta z \approx \Delta x_1 F_{x_1}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) + \Delta x_2 F_{x_2}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) + \ldots + \Delta x_n F_{x_n}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$$

$$\Delta_z \approx \Delta_{x_1} \left|F_{x_1}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)\right| + \Delta_{x_2} \left|F_{x_2}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)\right| + \ldots + \Delta_{x_n} \left|F_{x_n}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)\right|$$

$$|\Delta z| \le \Delta_z$$

Example

$$z = F(x,y) = \sqrt{1 + x^2 + y} \qquad x = 1.0 \pm 0.1 \qquad y = 2.0 \pm 0.5$$

$$\tilde{x} = 1.0 \quad \Delta_x = 0.1 \qquad \tilde{y} = 2.0 \quad \Delta_y = 0.5$$

$$F_x(x,y) = \frac{x}{\sqrt{1 + x^2 + y}} \qquad F_y(x,y) = \frac{1}{2\sqrt{1 + x^2 + y}}$$

$$F_x(\tilde{x},\tilde{y}) = \frac{1}{\sqrt{1 + 1^2 + 2}} = 0.5 \qquad F_y(\tilde{x},\tilde{y}) = \frac{1}{2\sqrt{1 + 1^2 + 2}} = 0.25$$

$$\Delta_z \approx \Delta_x \cdot \left|F_x(\tilde{x},\tilde{y})\right| + \Delta_y \cdot \left|F_y(\tilde{x},\tilde{y})\right| = 0.1 \times 0.5 + 0.5 \times 0.25 = 0.175$$

$$\delta_z \approx \Delta_z / F(\tilde{x},\tilde{y}) = 0.175/2 = 0.0875$$

Perturbation Experiment 1D

$$\tilde{x} = x + \Delta x \longrightarrow \boxed{F(x)} \longrightarrow \tilde{y} = y + \Delta y$$

- $F(x)$ is often:
  - Complicated, expensive, external code
- $\tilde{x} = x + \Delta x$ is input data:
  - Initial value, physical constant, problem parameter
- $\tilde{y} = y + \Delta y$ is output data:
  - Result, numbers, arrays

$$\tilde{y} = F(\tilde{x}) \qquad y_{\text{exp}} = F(\tilde{x} + \Delta_x) \longrightarrow \Delta_y \approx \left| y_{\text{exp}} - \tilde{y} \right|$$

Perturbation Experiment 2D

$$z = F(x, y)$$

1) Best guess $\quad \tilde{z} = F(\tilde{x}, \tilde{y})$

2) Exp1 $\quad z_{1,\exp} = F(\tilde{x} + \Delta_x, \tilde{y})$

3) Exp2 $\quad z_{2,\exp} = F(\tilde{x}, \tilde{y} + \Delta_y)$

4) Sum up $\quad E_z \approx \left| z_{1,\exp} - \tilde{z} \right| + \left| z_{2,\exp} - \tilde{z} \right|$

Taylor's Theorem

$$f \in C^n[a,b] \ \& \ f^{(n+1)} \exists \text{ on } (a,b) \quad \forall x, c \in [a,b]$$

$$\exists \xi \,\middle|\, \xi \text{ is between } x \text{ and } c$$

$$f(x) = f(c) + \sum_{k=1}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k + E_n(x)$$

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1}$$

Example

$$f(x) = \ln x; \qquad a = 1,\ b = 2,\ c = a \qquad\qquad f^{(k)}(x) = (-1)^{k-1}(k-1)!\,x^{-k}$$

$$\ln x = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \ldots + (-1)^{n-1}\frac{1}{n}(x-1)^n + E_n(x)$$

$$E_n(x) = \frac{1}{\xi^n}\frac{(-1)^n}{(n+1)}(x-1)^{n+1}; \quad 1 < \xi < x \longrightarrow \left|E_n(x)\right| \le \frac{1}{n+1}(x-1)^{n+1}$$

Estimating Accuracy

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + (-1)^{n-1}\frac{1}{n} + E_n(2)$$

$$\left|E_n(2)\right| \le \frac{1}{n+1} \le 10^{-6} \longrightarrow n+1 \ge 10^{6}$$

Another Form

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{1}{k!} f^{(k)}(x) h^k + E_n(h)$$

$$E_n(h) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) h^{n+1} \quad \xi \text{ is between } x \text{ and } x+h$$

$$E_n(h) = \frac{1}{(n+1)!} f^{(n+1)}(x+\theta h) h^{n+1} \qquad 0 < \theta < 1$$

Taylor's Theorem in 2D

$$f:\mathbb{R}^2 \to \mathbb{R}$$

$$\begin{aligned} f(a+h,b+k) = {} & f(a,b) + \\ & + h\frac{\partial f(a,b)}{\partial x} + k\frac{\partial f(a,b)}{\partial y} + \\ & + \frac{1}{2}h^2\frac{\partial^2 f(a,b)}{\partial x^2} + hk\frac{\partial^2 f(a,b)}{\partial x \partial y} + \frac{1}{2}k^2\frac{\partial^2 f(a,b)}{\partial y^2} + \\ & + E_2(h,k) \end{aligned}$$

General 2D Form

$$f : \mathbb{R}^2 \to \mathbb{R}$$

$$f(a+h, b+k) = \sum_{i=0}^{n} \frac{1}{i!} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^i f(a,b) + E_n(h,k)$$

$$E_n(h,k) = \frac{1}{(n+1)!} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^{n+1} f(a+\theta h, b+\theta k) \qquad 0 < \theta < 1$$

Binominal Theorem

$$(x+y)^n = x^n + nx^{n-1}y + \ldots \binom{n}{k} x^{n-k} y^k + \ldots + nxy^{n-1} + y^n$$

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}; \quad \binom{n}{k} = \binom{n}{n-k}; \quad \binom{n}{0} = 1; \quad \binom{n}{1} = n.$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Big O

$$\{x_n\},\ \{\alpha_n\}$$

$$x_n = O(\alpha_n) \quad \exists C, N: \quad |x_n| \le C|\alpha_n| \quad \forall n \ge N$$

$$\left|\frac{x_n}{\alpha_n}\right| \le C \quad \text{when} \quad n \to \infty$$

Little o

$$\{x_n\},\ \{\alpha_n\}$$

$$x_n = o(\alpha_n) \qquad \lim_{n\to\infty}\frac{x_n}{\alpha_n} = 0$$

$$\frac{n+1}{n^2} = O\left(\frac{1}{n}\right) \qquad \frac{1}{n\ln n} = o\left(\frac{1}{n}\right) \qquad e^{-n} = o\left(\frac{1}{n^2}\right)$$

O Notation

$$\sin x = x - \frac{x^3}{6} + O\left(x^5\right) \qquad \left(x \to 0\right)$$

$$\left|\sin x - x + \frac{x^3}{6}\right| \le Cx^5 \qquad \text{in an neighbourhood of } 0$$

O Notation

$$f(x) = O\big(g(x)\big) \qquad \big(x \to 0, \quad x \to x_0, \quad x \to \infty\big)$$

$$\left|f(x)\right| \le C\left|g(x)\right|$$

$$f(x) = o\big(g(x)\big) \quad \leftrightarrow \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

Orders of Convergence

$$x_n \xrightarrow[n\to\infty]{} L$$

At least linear

$$|x_{n+1} - L| \le c\,|x_n - L| \qquad (c < 1, \quad n \ge N)$$

Superlinear

$$|x_{n+1} - L| \le \varepsilon_n\,|x_n - L| \qquad (\varepsilon_n \to 0)$$

At least quadratic

$$|x_{n+1} - L| \le c\,|x_n - L|^2$$

Polynomials

$$p = p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \qquad a_n \neq 0$$

$$\deg p(x) \equiv n$$

$$\Pi_n \equiv \left\{ p \middle| \deg p \leq n \right\}$$

$$\forall p, q \in \Pi_n \longrightarrow p + q \in \Pi_n \quad \& \quad \lambda p \in \Pi_n$$

$$p(x) = d(x) \cdot q(x) + r(x) \quad \deg q < \deg p \qquad \deg r < \deg p$$

Horner's Method

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = \sum_{k=0}^{n} \left(a_k \prod_{j=1}^{k} x \right)$$

$$p(x) = \left(\left(\left(\ldots \left(\left(x a_n + a_{n-1} \right) x + a_{n-2} \right) x + \ldots + a_3 \right) x + a_2 \right) x + a_1 \right) x + a_0$$

```
p = a(n);
for k = n-1:-1:0
   p = p*x + a(k);
end
```

```
p = a[n]
for k in range(n-1,-1,-1):
   p = p*x + a[k]
```

Important

- Absolute/Relative Error
- Error Sources
- Significant Digits
- Propagation of Errors
- Linearization/Taylor's Theorem
- Taylor's Theorem in 2D
- Order of Convergence