Total Error (contd.)

Recall Arithmetic Using n-Digit Rounding and Chopping

The computed value $fl(fl(x) \odot fl(y))$ involves an error which comprises of

- Error in $fl(x)$ and $fl(y)$ due to n-digit rounding or chopping.
- Error in $fl(fl(x) \odot fl(y))$ due to n-digit rounding or chopping.

The total error is defined as

$$(x \odot y) - fl(fl(x) \odot fl(y)) = [(x \odot y) - (fl(x) \odot fl(y))] + [(fl(x) \odot fl(y)) - fl(fl(x) \odot fl(y))]$$

in which the first term on the right hand side is called the propagated error and the second term is called the floating-point error.

Total Error (contd.)

Example:

Consider evaluating the integral

$$I_n = \int_0^1 \frac{x^n}{x+5}\,dx, \quad \text{for} \quad n = 0, 1, \cdots, 20.$$

The value of I_n can be obtained in two different iterative processes, namely,

- $I_n = \frac{1}{n} - 5I_{n-1}$, $I_0 = \ln(6/5)$ (called forward iteration) and
- $I_{n-1} = \frac{1}{5n} - \frac{1}{5}I_n$, $I_{30} = 0.54046330 \times 10^{-2}$ (called backward iteration).

Total Error (contd.)

Example: The following table shows the computed value of I_n using both iterative formulas along with the exact value. The numbers are rounded to 6-digits.

n	Forward Iteration	Backward Iteration	Exact Value
1	0.088392	0.088392	0.088392
5	0.028468	0.028468	0.028468
10	0.015368	0.015368	0.015368
15	0.010522	0.010521	0.010521
20	0.004243	0.007998	0.007998
25	11.740469	0.006450	0.006450
30	-36668.803026	Not Computed	0.005405

Condition Number

For a given function $f: \mathbb{R} \to \mathbb{R}$, consider evaluating $f(x)$ at an approximate value x_A rather than at x

The question is **how well does $f(x_A)$ approximate $f(x)$?**

Using the mean-value theorem, we get

$$f(x) - f(x_A) = f'(\xi)(x - x_A),$$

where ξ is an unknown point between x and x_A.

The relative error of $f(x)$ with respect to $f(x_A)$ is given by

$$E_r(f(x_A)) = \frac{f'(\xi)}{f(x)}(x - x_A) = \left(\frac{f'(\xi)}{f(x)}x\right) E_r(x_A).$$

Condition Number (contd.)

Since x_A and x are assumed to be very close to each other and ξ lies between x and x_A, we may make the approximation

$$f(x) - f(x_A) \approx f'(x)(x - x_A).$$

Using this, we have

$$E_r(f(x_A)) \approx \left(\frac{f'(x)}{f(x)}x\right) E_r(x_A).$$

Definition (Condition number of a function)

The condition number of a continuously differentiable function f at a point $x = c$ is given by

$$\left|\frac{f'(c)}{f(c)}c\right|$$

Condition Number (contd.)

Definition (Well-Conditioned and Ill-Conditioned)

The process of evaluating a continuously differentiable function f at a point $x = c$ is said to be **well-conditioned** if the condition number

$$\left|\frac{f'(c)}{f(c)}c\right|$$

at c is small.

The process of evaluating a function at $x = c$ is said to be **ill-conditioned** if it is not well-conditioned.

"How small the condition number should be?"

Condition Number (contd.)

Example: Consider the function $f(x) = \sqrt{x}$, for all $x \in [0, \infty)$. Then

$$f'(x) = \frac{1}{2\sqrt{x}}, \text{ for all } x \in [0, \infty).$$

The condition number of f is

$$\left|\frac{f'(x)}{f(x)}x\right| = \frac{1}{2}, \text{ for all } x \in [0, \infty).$$

Thus, we have

$$|E_r(f(x_A))| \approx \frac{1}{2}|E_r(x_A)|.$$

Condition Number (contd.)

Example: Consider the function

$$f(x) = \frac{10}{1 - x^2}, \text{ for all } x \in \mathbb{R}.$$

Then $f'(x) = 20x/(1 - x^2)^2$, so that

$$\left| \frac{f'(x)}{f(x)} x \right| = \left| \frac{(20x/(1 - x^2)^2)x}{10/(1 - x^2)} \right| = \frac{2x^2}{|1 - x^2|}$$

and this number can be quite large for $|x|$ near 1.

Condition Number (contd.)

Example: Consider the function

$$f(x) = \sqrt{x+1} - \sqrt{x}, \text{ for all } x \in (0, \infty).$$

For a sufficiently large x, the condition number of this function is

$$\left|\frac{f'(x)}{f(x)}x\right| = \frac{1}{2}\left|\frac{\left(\frac{1}{\sqrt{x+1}} - \frac{1}{\sqrt{x}}\right)}{\sqrt{x+1} - \sqrt{x}}x\right| = \frac{1}{2}\frac{x}{\sqrt{x+1}\sqrt{x}} \leq \frac{1}{2},$$

which is quite good.

Stable and Unstable Computations

Definition (Stability and Instability in Evaluating a Function)

Suppose there are n steps to evaluate a function $f(x)$ at a point $x = c$. Then the total process of evaluating this function is said to have **instability** if atleast one of the n steps is ill-conditioned. If all the steps are well-conditioned, then the process is said to be **stable**.

Stable and Unstable Computations (contd.)

Let us analyze the computational process of the function

$$f(x) = \sqrt{x+1} - \sqrt{x}$$

x_0

The computational process consists of the following four computational steps:

$$f_4(t) = t - x_3 \qquad \begin{cases} x_0 \Rightarrow \text{given number}, \\ x_1 := x_0 + 1, \\ x_2 := \sqrt{x_1}, \\ x_3 := \sqrt{x_0}. \\ x_4 := x_2 - x_3. \end{cases}$$

$f_1(t) = t + 1$

$f_2(t) = \sqrt{t}$

$f_3(t) = \sqrt{t}$

Stable and Unstable Computations (contd.)

Now consider the last two steps where we already computed x_2 and now going to compute x_3 and finally evaluate the function

$$f_4(t) := x_2 - t.$$

At this step, the condition number for f_4 is given by

$$\left|\frac{f_4'(t)}{f_4(t)} t\right| = \left|\frac{t}{x_2 - t}\right|.$$

Thus, f_4 is ill-conditioned when t approaches x_2.

Stable and Unstable Computations (contd.)

Let us rewrite the same function $f(x)$ as $\tilde{f}(x) = \dfrac{1}{\underbrace{\sqrt{x+1}+\sqrt{x}}}$.

x_0

$x_1 \quad x_0 + 1$

$x_2 \quad \sqrt{x_1}$

$x_3 \quad \sqrt{x_0}$

$x_4 = \underbrace{x_2 + x_3}$

$x_5 = 1/x_4$

Linear System: General Form

General form of a system of n linear equations in n variables is

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\cdots \\
&\cdots \\
&\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}
$$

General Form of Linear System (contd.)

These equations can be written in the matrix notation as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

The last equation is usually written in the following short form

$$Ax = b,$$

where

- A stands for the $n \times n$ matrix with entries a_{ij}.
- $x = (x_1, x_2, \cdots, x_n)^T$

General Form of Linear System (contd.)

Let us now state a result concerning the solvability of the system

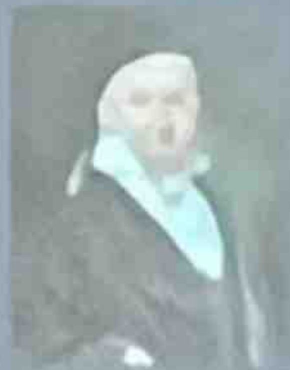
$$Ax = b.$$

Theorem

Let A be an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$. Then the following statements concerning the system of linear equations $A\mathbf{x} = \mathbf{b}$ are equivalent.

1. $\det(A) \neq 0$
2. For each right hand side vector $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x}$
3. For $\mathbf{b} = 0$, the only solution of the system $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = 0$

Linear Systems: Naive Gaussian Elimination Method



Carl Friedrich Gauss (1777–1855) German mathematician

Consider the following system:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \quad \Longleftrightarrow \quad \text{Upper triangular System} \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3.
\end{aligned}
$$