§ Rounding Errors

Round-off errors are unavoidable

and accumulate as computations go on.

$E_n \equiv$ magnitude of rounding error after n subsequent operations

* linear growth : $E_n \approx CnE_0$ for some constant C

Usually unavoidable but acceptable as long as C and E_0 are sufficiently small.

* exponential growth: $E_n \approx C^n E_0$ for some constant $C > 1$

Overflow!

example: compute the series $P_n = \frac{1}{3^n}$ with single-precision real numbers

Method 1

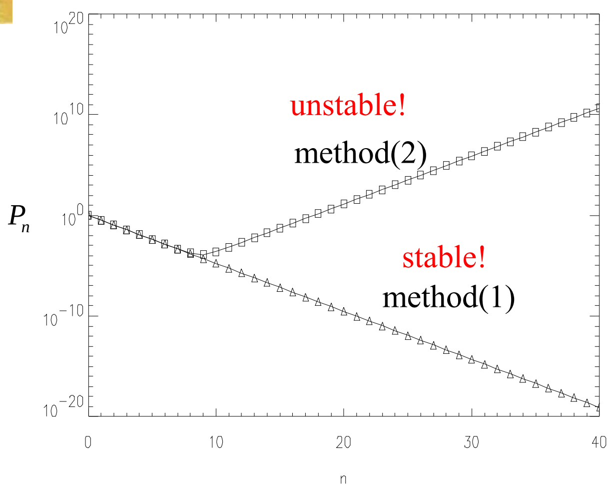
```
P(0)=1
DO n=1,100
    P(n)=1./3.*P(n-1)
END DO
```

Method 2

```
P(0)=1
P(1)=1./3.
DO n=2,100
    P(n)=10./3.*P(n-1)–P(n-2)
END DO
```

Method 1: $fl(P_n) = fl\left(fl\left(\frac{1}{3}\right) * fl(P_{n-1}) \right)$

Method 2: $fl(P_n) = fl\left(fl\left(\frac{10}{3}\right) * fl(P_{n-1}) - fl(P_{n-2}) \right)$



Method 1: $fl(P_n) = fl\left(fl\left(\frac{1}{3}\right) * fl(P_{n-1}) \right)$

Method 2: $fl(P_n) = fl\left(fl\left(\frac{10}{3}\right) * fl(P_{n-1}) - fl(P_{n-2}) \right)$

Method 2

```
P(0)=1
P(1)=1./3.
DO n=2,100
    P(n)=10./3.*P(n-1)–P(n-2)
END DO
```

$P_n = \frac{1}{3^n}$

formula correct!

$$P_n = \rho^n$$

$$\rho^n = \frac{10}{3}\rho^{n-1} - \rho^{n-2}$$

$$\rho^2 - \frac{10}{3}\rho + 1 = 0$$

$$(\rho - 3)\left(\rho - \frac{1}{3}\right) = 0$$

$$P_n = A \cdot 3^n + B \cdot \frac{1}{3^n}$$

$$\left.\begin{array}{l} P_0 = 1 = A + B \\ P_1 = \frac{1}{3} = 3A + \frac{B}{3} \end{array}\right\} \Rightarrow \begin{cases} A = 0 \\ B = 1 \end{cases}$$

Ways of Avoiding Rounding Errors:

① **Reduce # of computations as many as possible.**

$$\pi + e = 3.141592653... + 2.71828182... = 5.85987448...$$

$$\pi * e = 3.141592653... * 2.71828182... = 8.53973422...$$

7 digits + rounding method:

$$fl(fl(\pi) + fl(e)) = fl(3.141593 + 2.718282) = 5.859875$$

$$fl(fl(\pi) * fl(e)) = fl(3.141593 * 2.718282)$$
$$= fl(8.539735703...) = 8.539736$$

② **Avoid substraction of two nearly equal numbers.**

$$fl(x) - fl(y) = 0.3141593 \times 10^1 - 0.3141291 \times 10^1 = 0.3020000 \times 10^{-3}$$

~ lose 4 digits of significance

(Any further calculations can have only 3, instead of 7, digits of significance.)

③ **Avoid dividing by a small number.**

original rounding error $= \delta$ exact number $= z = fl(z) + \delta$

divided by a small number $\varepsilon = 10^{-6}$

$$\text{rounding error} = \left| \frac{z}{\varepsilon} - \frac{fl(z)}{\varepsilon} \right| = \left| \frac{\delta}{\varepsilon} \right| = 10^6 |\delta|$$