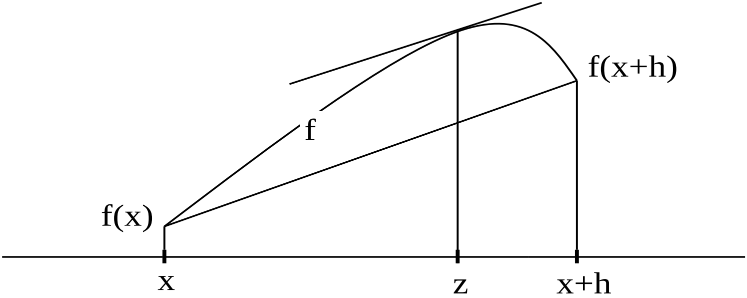
Taylor Series, Derivative Approximation, and Numerical Cancellation

- 1 lecture
- References:
  - Overton, Chapter 11
  - Cheney & Kincaid, Sections 1.1, 1.4, 4.3

The **mean value theorem** (**MVT**):
Let $f(x)$ be differentiable. For **some** z in $[x, x+h]$:

$$f'(z) = \frac{f(x+h) - f(x)}{h}.$$

This is *intuitively* clear from:



The Mean Value Theorem

We can **rewrite** the MV formula as:

$$f(x+h) = f(x) + hf'(z).$$

A generalization if f is **twice** differentiable is

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(z),$$

for some z in $[x, x+h]$.

The Taylor Series

Taylor Theorem:

$$\boxed{f(x+h) = \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!} h^k + E_{n+1}}$$

where the error (remainder)

$$E_{n+1} = \frac{f^{(n+1)}(z)}{(n+1)!} h^{n+1}, \quad z \in [x, x+h]$$

Taylor Theorem:

$$\boxed{f(x+h) = \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!} h^k + E_{n+1}}$$

where the error (remainder)

$$E_{n+1} = \frac{f^{(n+1)}(z)}{(n+1)!} h^{n+1}, \quad z \in [x, x+h]$$

Taylor series.

$$\boxed{f(x+h) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x)}{k!} h^k, \quad |h| < R}$$

The RHS is assumed to converge with radius of convergence R.

Example: For $f(x) = \sin x$,

$$\begin{aligned} \sin(x+h) &= \sin(x) + \sin'(x)h + \frac{\sin''(x)}{2!}h^2 \\ &+ \frac{\sin'''(x)}{3!}h^3 + \frac{\sin''''(x)}{4!}h^4 + \cdots, \quad |h| < \infty \end{aligned}$$

Letting $x = 0$, we get

$$\sin(h) = h - \frac{1}{3!}\,h^3 + \frac{1}{5!}\,h^5 - \cdots,$$

since **even** derivatives of $\sin x$ are zero at $x = 0$, and **odd** ones are ± 1.

Numerical Approximation to $f'(x)$

If h is **small**, $f'(x)$ is nearly the **slope** of the line through $(x, f(x))$ and $(x+h, f(x+h))$. We write

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}, \quad \text{forward difference.}$$

Numerical Approximation to $f'(x)$

If h is **small**, $f'(x)$ is nearly the **slope** of the line through $(x, f(x))$ and $(x+h, f(x+h))$. We write

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}, \quad \text{forward difference.}$$

How **good** is this approximation?

Numerical Approximation to $f'(x)$

If h is **small**, $f'(x)$ is nearly the **slope** of the line through $(x, f(x))$ and $(x+h, f(x+h))$. We write

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}, \quad \text{forward difference.}$$

How **good** is this approximation?

By the **Taylor Theorem**:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(z).$$

Then

$$\frac{f(x+h) - f(x)}{h} - f'(x) = \frac{h}{2} f''(z),$$

which is called the **discretization error**, the difference between what we want and our approximation, using the **discretization size** h. We say the <u>discretization error is $O(h)$</u>.

Computing the Approximation to $f'(x)$

Approximate the derivative of $f(x) = \sin(x)$ at $x = 1$. Compute the **exact** discretization errors for $h = 10^{-1}, \ldots, 10^{-20}$.

```
int main()
{int n; double x,h,approx,exact,error;
 x = 1.0; h = 1.0; n = 0;
 printf("\n h exact approx error");
 while(n<20) {
   n++;
   h = h/10;       /* h=10^(-n) */
   approx = (sin(x+h)-sin(x))/h;       /*app.deriv.*/
   exact = cos(x);        /*exact derivative */
   error = approx - exact;     /*disczn error */
   printf("... \n",h,approx,exact,error);
 }
}
```

Convergence of Approximation

h	approx	exact	error
1.0e-03	5.398815e-01	5.403023e-01	-4.208255e-04
1.0e-04	5.402602e-01	5.403023e-01	-4.207445e-05
1.0e-05	5.402981e-01	5.403023e-01	-4.207362e-06
1.0e-06	5.403019e-01	5.403023e-01	-4.207468e-07
1.0e-07	5.403023e-01	5.403023e-01	-4.182769e-08
1.0e-08	5.403023e-01	5.403023e-01	-1.407212e-08
1.0e-09	5.403024e-01	5.403023e-01	5.254127e-08
1.0e-10	5.403022e-01	5.403023e-01	-5.848104e-08
1.0e-11	5.403011e-01	5.403023e-01	-1.168704e-06
1.0e-12	5.403455e-01	5.403023e-01	4.324022e-05
1.0e-13	5.395684e-01	5.403023e-01	-7.339159e-04
1.0e-14	5.329071e-01	5.403023e-01	-7.395254e-03
1.0e-15	5.551115e-01	5.403023e-01	1.480921e-02
1.0e-16	0.000000e+00	5.403023e-01	-5.403023e-01

Convergence of Approximation, ctd

- When h changes from 10^{-3} to 10^{-8}, the approximation gets **better**, and when h is reduced by 10, the **discretization error** is reduced by ~ 10, so the error **is** $O(h)$.
- When $h = 10^{-9}$, the approximation starts to get **worse!**
- When h changes from 10^{-9} to 10^{-16}, the approximation gets **worse** and **worse**.
- When $h = 10^{-16}$, `approx` becomes 0.

Q: Why **??**

Explanation of Accuracy Loss

- If $x = 1$, and $h \leq \frac{1}{2}\epsilon \approx 1.1 \times 10^{-16}$, $x + h$ has the **same numerical value** as x, so $f(x + h)$ and $f(x)$ **cancel** to give **0** and the quantity `approx` has **no digits of precision**.

Explanation of Accuracy Loss

- If $x = 1$, and $h \leq \frac{1}{2}\epsilon \approx 1.1 \times 10^{-16}$, $x + h$ has the **same numerical value** as x, so $f(x + h)$ and $f(x)$ **cancel** to give **0** and the quantity `approx` has **no digits of precision**.
- When h is a **little** bigger than $\epsilon/2$, the values **partially cancel**. For example, suppose that the first 10 digits of $\sin(x + h)$ and $\sin(x)$ are the same. Then, even though $\sin(x + h)$ and $\sin(x)$ are **accurate to 16 digits**, the **difference** has only **6 accurate digits**.

Explanation of Accuracy Loss

- If $x = 1$, and $h \leq \frac{1}{2}\epsilon \approx 1.1 \times 10^{-16}$, $x + h$ has the **same numerical value** as x, so $f(x + h)$ and $f(x)$ **cancel** to give **0** and the quantity `approx` has **no digits of precision**.
- When h is a **little** bigger than $\epsilon/2$, the values **partially cancel**. For example, suppose that the first 10 digits of $\sin(x + h)$ and $\sin(x)$ are the same. Then, even though $\sin(x + h)$ and $\sin(x)$ are **accurate to 16 digits**, the **difference** has only **6 accurate digits**.
- In summary, using h **too big** means a big **discretization** error, while using h **too small** means a big **cancellation** error.

  For the function $f(x) = \sin(x)$, at $x = 1$, the best choice of h is about 10^{-8}, or $\sim \sqrt{\epsilon}$.

Numerical Cancellation

The cancellation phenomenon occurs when we do a subtraction of two **nearly equal** numbers and it is one of the main causes for deterioration in accuracy.

The cancellation phenomenon occurs when we do a subtraction of two **nearly equal** numbers and it is one of the main causes for deterioration in accuracy.

Theoretical Analysis

In a computation, usually operands have some errors.
Instead of the correct values x and y, the computer works with two perturbed floating point numbers:

$$\hat{x} = x(1+\delta_1), \qquad \hat{y} = y(1+\delta_2),$$

where the errors δ_1 and δ_2 may be due to previous computations, physical experiments and/or rounding.

Suppose we want to compute $x - y$. But we can only compute $\hat{x} - \hat{y}$. The computed value of $\hat{x} - \hat{y}$ is

$$(\hat{x} - \hat{y})(1+\delta_3), \quad |\delta_3| < \epsilon$$

Numerical Cancellation Error

Is $(\hat{x}-\hat{y})(1+\delta_3)$ a good approximation to $x-y$?

The relative error:

$$
\begin{aligned}
&\left|\frac{(\hat{x}-\hat{y})(1+\delta_3)-(x-y)}{x-y}\right| \\
&= \left|\frac{x(1+\delta_1)(1+\delta_3)-y(1+\delta_2)(1+\delta_3)-(x-y)}{x-y}\right| \\
&= \left|\delta_3+\frac{x}{x-y}\delta_1+\frac{x}{x-y}\delta_1\delta_3+\frac{y}{x-y}\delta_2+\frac{y}{x-y}\delta_2\delta_3\right|.
\end{aligned}
$$

This suggests that when

$$|x-y| \ll |x|, |y|,$$

it is very likely that the relative error is very large even if $|\delta_1|$ and $|\delta_2|$ are very small and $\delta_3 = 0$.

In numerical computating, avoid numerical cancellation if possible.

How to solve $ax^2 + bx + c = 0$ in a reliable way?

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

Given $a = 1$, $b = -10000$, $c = 1$, the exact solutions are True solutions (accurate up to the last digits)

$$x_1 = 9999.999899999998, \quad x_2 = 0.000100000000100000002$$

In single precision, the formulas give

$$x_1 = 10,000.0, \text{ very good}, \quad x_2 = 0, \text{ completely wrong}$$

How to solve $ax^2 + bx + c = 0$ in a reliable way?

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

Given $a = 1$, $b = -10000$, $c = 1$, the exact solutions are True solutions (accurate up to the last digits)

$$x_1 = 9999.9998999999998, \qquad x_2 = 0.000100000000100000002$$

In single precision, the formulas give

$$x_1 = 10,000.0, \text{ very good}, \qquad x_2 = 0, \text{ completely wrong}$$

Reason: $\sqrt{b^2 - 4ac} \approx -b$, there is a numerical cancellation in computing $-b - \sqrt{b^2 - 4ac}$.

How to avoid the problem?

How to solve $ax^2 + bx + c = 0$ in a reliable way?

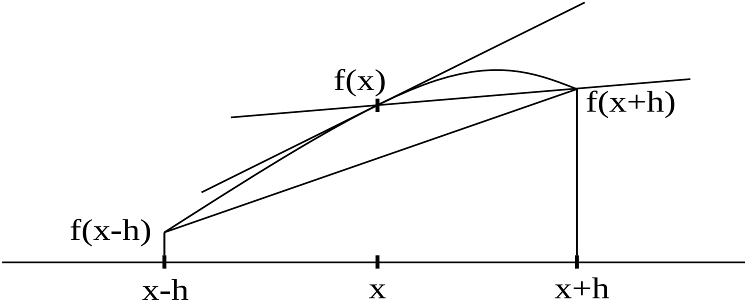
How to avoid the problem?

Idea: Rationalization

$$
\begin{aligned}
x_2 &= \frac{(-b - \sqrt{b^2 - 4ac})(-b + \sqrt{b^2 - 4ac})}{2a(-b + \sqrt{b^2 - 4ac})} \\
&= \frac{2c}{-b + \sqrt{b^2 - 4ac}} \\
&= \frac{c}{ax_1}
\end{aligned}
$$

Using the above formula, the computed $x_2 = 10^{-4}$, much more accurate.

More Accurate Numerical Differentiation



As h **decreases**, the line through $(x-h, f(x-h))$ and $(x+h, f(x+h))$ gives a **better approximation** to the **slope of the tangent** to f at x than the line through $(x, f(x))$ and $(x+h, f(x+h))$.

Central Difference Approximation

This observation leads to the approximation:

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h},$$

the **central difference** formula.

Analyzing Central Difference Formula

From the **Taylor Theorem:**

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(z_1),$$
$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(z_2),$$

with z_1 between x and $x+h$ and z_2 between x and $x-h$.

From the **Taylor Theorem:**

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(z_1),$$
$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(z_2),$$

with z_1 between x and $x+h$ and z_2 between x and $x-h$.

Subtracting the 2nd from the 1st:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{h^2}{12}(f'''(z_1) + f'''(z_2))$$

Discretization error $\frac{h^2}{12}(f'''(z_1) + f'''(z_2))$ is $O(h^2)$ **instead of** $O(h)$.