Numerical Derivative

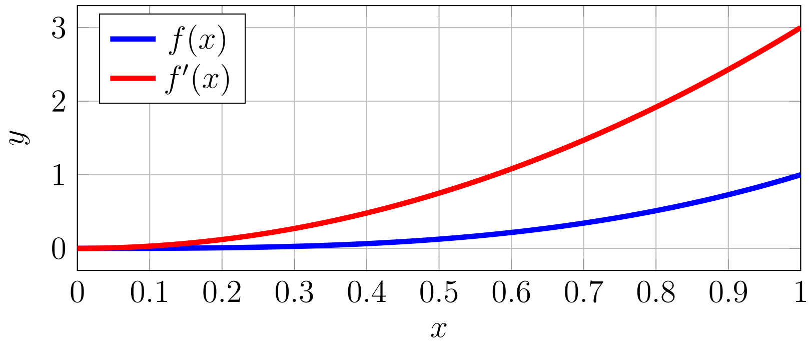
Let us consider the following equation

$$f(x) = x^3$$

From calculus we know that the derivate of the function is given by

$$f'(x) = 3x^2$$

Plotting them together will give us a function that looks like



So, if we want to know the value of the derivative at $x = 0.5$ all we need to do is plug in to $3x^2$. What we will get is

$$f'(0.5) = 3 \times 0.5^2 = 0.75$$

Teaching computer to differential symbolically like what we did above is actually quite complicated[1]. Luckily, numerically differentiate is quite easy.

First Attempt

Remember our goal is to calculate the derivative at $x = 0.5$ of the function $f(x) = x^3$.

The most naive way to calculate derivative is to go back the definition of derivative

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

The $\lim\limits_{h \to 0}$ part means that we need to evalate the expression when h is really close but not equal to zero. Of course, computer can't do that.

But, we can tell computer to "sort of" do it. If we let h to be a small number, then the expression would be really close to the derivative

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \tag{1}$$

[1] See `https://en.wikipedia.org/wiki/Automatic_differentiation`

This method is called *forward difference*.

Let us try this idea. Let $h = 0.1$. Then, the derivative at $x = 0.5$ is approximately

$$f'(0.5) \approx \frac{f(0.5+0.1) - f(0.5)}{0.1} = \frac{0.6^3 - 0.5^3}{0.1} = \frac{0.216 - 0.125}{0.1} = 0.91$$

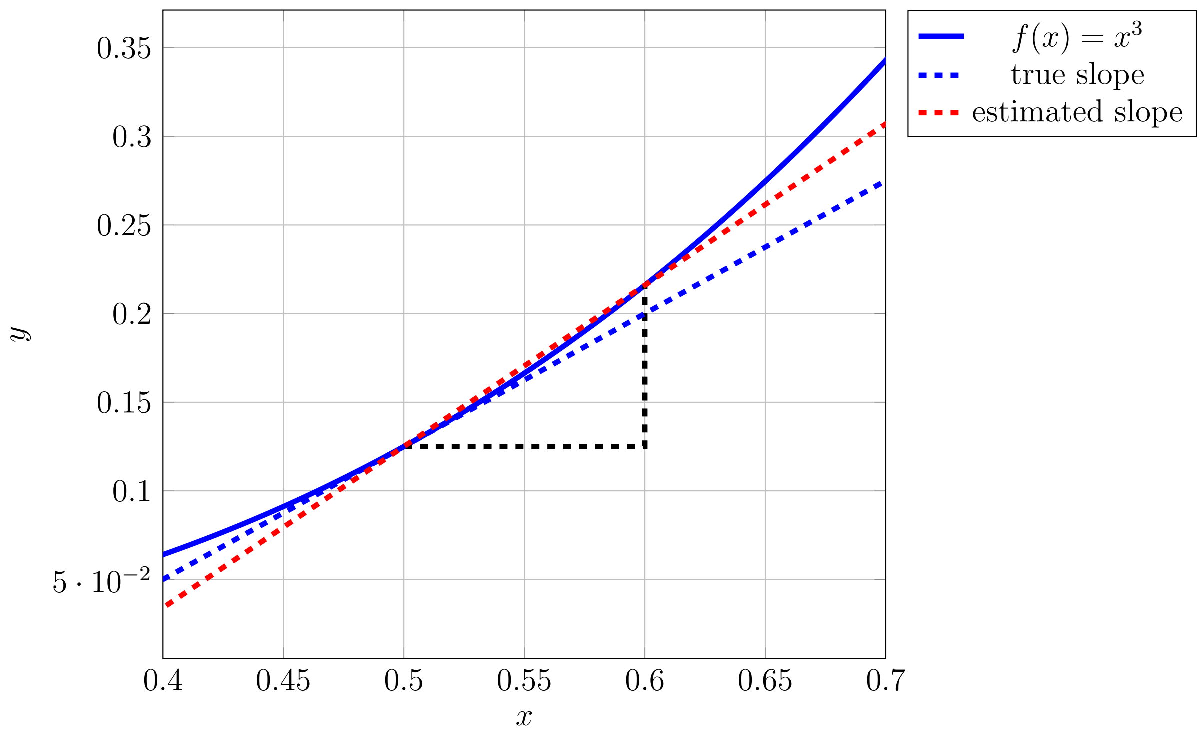
Not exactly close to the real answer

Let us try a smaller $h = 0.0001$

$$f'(0.5) \approx \frac{f(0.5+0.0001) - f(0.5)}{0.1} = \frac{0.5001^3 - 0.5^3}{0.1} = \frac{0.216 - 0.125}{0.1} = 0.75015$$

much closer than what we had above.

You can see from the two example above that if you want accurate answer h needs to be really small. The reason can be illustrated with the following picture.



The triangle that we use is too big and it's not even centered at the point we want to calculate.

Second attempt

Let's do an easy fix by at least center the triangle at the point we want to calculate. So we have

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

This method is called centered finite difference.

Let us do that using the same bit width as we use last time

$$f'(x) \approx \frac{f(0.55) - f(0.45)}{0.1} = 0.7525$$

Much much better than last time.

If we use the small bound we have last time

$$f'(x) \approx \frac{f(0.50005) - f(0.49995)}{0.1} = 0.75$$

It's 0.75 plus something small that is beyond the precision of our computer. This is done by just shifting the triangle to the center.

Why is it such a big effect

We will now try to understand why it is much more precise. Let us now derive the Equation 1 from Taylor series. Recall that

$$f(x+h) = f(x) + f'(x)h + \frac{f''(\xi)}{2}h^2$$

where $\xi \in [x, x+h]$. We can obtain Equation 1 by moving $f(x)$ to the other side and obtain

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

The most important term here is the error term ($\mathcal{E}$):

$$\mathcal{E}(h) = \frac{f''(\xi)}{2}h = \mathcal{O}(h)$$

This is linear in h. First, the smaller the h is the more accurate our result would be. Second, if we treat $f''(\xi)$ as a constant[2]. This means that if our h is smaller by the factor of 10. The error will also be smaller by the factor of 10.

Now let us try applying the same approach for Equation 2. Let us first compute the right hand side of the equation 1.

Let us modify the right hand side 2 a little bit by just calling $h \to 2h$ to get rid of all half h.

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h} \tag{3}$$

Let us compute the right hand side of the above expression

$$\begin{aligned} f(x+h) =& f(x) + f'(x)h + \frac{f''(x)}{2}h^2 + \frac{f'''(\xi^+)}{3!}h^3 \\ f(x-h) =& f(x) - f'(x)h + \frac{f''(x)}{2}h^2 - \frac{f'''(\xi^-)}{3!}h^3 \end{aligned}$$

Note first that the ξ in the top and the bottom equation may not be the same ξ. The $\xi^+ \in [x, x+h]$ and $\xi^- \in [x-h, x]$.

[2]It actually changes a bit with h but the bound on $f''(h)$ is, for all practial intent and purpose, constant.

Subtracting the two gives

$$f(x+h) - f(x-h) = 2f'(x)h + \frac{f'''(\xi^+) + f'''(\xi^-)}{3!}h^3.$$

Note that the h^2 term cancels magically. Dividing my $2h$ on both sides yields the formula we want

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{f'''(\xi^+) + f'''(\xi^-)}{2}\frac{h^2}{3!}$$

Recall intermediate value theorem that if $f'''(x)$ is continuous from $[\xi^+, \xi^-]$ then there is $\xi \in [\xi^+, \xi^-]$ such that it attains the average value:

$$f'''(\xi) = \frac{f'''(\xi^+) + f'''(\xi^-)}{2}$$

So we have

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + f'''(\xi)\frac{h^2}{3!} = f'(x) + \mathcal{O}(h^2)$$

This means that the error $\mathcal{E}(h)$ goes like h^2. This is very impressive since if we make h smaller by factor of 10 our result is more accurate by a factor of 100. This kind of quadratic convergence is usually what we seek for in numerical program since your effor will be magnified by a huge amount.

We can actually make the convergence even faster by including more terms. You will see some example in the homework. We use theses more accurate derivative only in some specialize application. For most application the centered finite difference is good enough.

Second Derivative

Sometime we are intested in finding the second derivative. We can again go back to the definition.

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

But this evaluation require you to know the $f'(x)$ which we actually don't have one.

Luckily even if we don't know $f'(x)$ we can approximate using what we did in the previous section. So let us use what we got in the previous section to approximate $f'(x+h)$ and $f'(x-h)$.

$$\begin{aligned} f'(x+h) \approx& \frac{f(x+h+h) - f(x+h-h)}{2h} = \frac{f(x+2h) - f(x)}{2h} \\ f'(x-h) \approx& \frac{f(x-h+h) - f(x-h-h)}{2h} = \frac{f(x) - f(x-2h)}{2h} \end{aligned}$$

Pluggin everything in Equation gives us

$$
\begin{aligned}
f''(x) \approx& \frac{f'(x+h) - f'(x-h)}{2h} \\
\approx& \frac{\left(\frac{f(x+2h) - f(x)}{2h}\right) - \left(\frac{f(x) - f(x-2h)}{2h}\right)}{2h} \\
\approx& \frac{f(x+2h) - 2f(x) + f(x-2h)}{4h^2}
\end{aligned}
$$

This is the same thing as (by calling $2h \to h$)

$$
f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \tag{4}
$$

This is also called centered finite difference for second derivative.

Let us try using this to calculate the second derivative of $f(x) = x^5$ at $x = 0.5$. We know that

$$
f''(x) = 20 \times x^3 \implies f''(0.5) = 20 \times 0.5^3 = 2.5
$$

Let us try to calculate the same number using Equation 4 with $h = 0.1$

$$
\begin{aligned}
f''(0.5) \approx& \frac{f(0.5+0.1) - 2f(0.5) + f(0.4)}{0.1^2} \\
\approx& \frac{0.6^5 - 2 \times 0.5^5 + 0.4^5}{0.1^2} \\
=& 2.55
\end{aligned}
$$

Let us try to calculate the same number using Equation 4 with $h = 0.01$

$$
\begin{aligned}
f''(0.5) \approx& \frac{f(0.5+0.01) - 2f(0.5) + f(0.5-0.01)}{0.01^2} \\
=& 2.5005
\end{aligned}
$$

Very close to what we had. We can do the same exercise as what we did to the first derivative to show that

$$
\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = f''(x) + \mathcal{O}(h^2)
$$

This is typically what we use in real world application.

Caveat. Fighting against rounding error.

In calculating derivative, it is tempting to makes h as small as possible in calculating derivative; for example why not just let $h = 10^{-30}$ when we calculate $f'(0.5)$. Let us try that

```
def f(x):
return x**3

x = 0.5
h = 1e-30

dfdx = (f(x+h) - f(x-h))/(2*h)
print dfdx
```

You will see that you get 0. Why?

Recall the exercise that I asked you in the first class that

```
x = 1e-10
(1+x) -1 == x
```

is false because of the fact that floating point can only keep finite number of significant digit. The same thing happens here $0.5 + 1e - 30$ is not representable using floating point. So it got truncated and everything is ruined.

The problem is actually a bit more subtle than that. If we let $h = 10^{-12}$, we would think that our answer would be extremely accurate. Let us considier the following example:

```
def f(x):
return 1000+x**3

x = 0.5
h = 1e-12
print x+h #0.500000000001

dfdx = (f(x+h) - f(x-h))/(2*h)
print dfdx #0.795807864051
```

Our answer is accurate to only 1 digit while we were expecting something of 20+ digit. The reason for this is that even though $x+h$ and $x-h$ are different $1+x^3$ are practically the same for $x+h$ and $x-h$. Whent the computer subtract the two numbers, only the last few significant digit are subtract off. The rest are thrown away. Then we multiply everything back by 10^{12} so a not a lot of information are left. So we get a bit rounding error.

The takehome lesson here is that if once h gets too small, rounding error will take over the truncation error. You can read more details in wikipedia on this issue `https://en.wikipedia.org/wiki/Numerical_differentiation`. This makes quadratic convergence very attractive since large h helps you avoid truncation error while the formula error is small by construction.