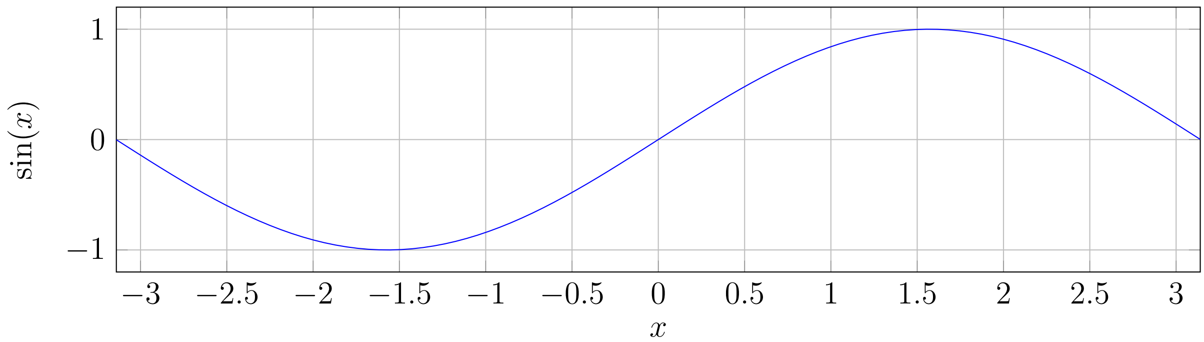
Numerical Methods

Taylor's Theorem

Last update: 2015/10/02 00:48:00

Let imagine that we are task to implement sin function using computer. In case you forgot this is what it looks like



The first approach you are going to do is something like this

```
import math
def sin_why_do_you_even_ask(x):
    return math.sin(x)
```

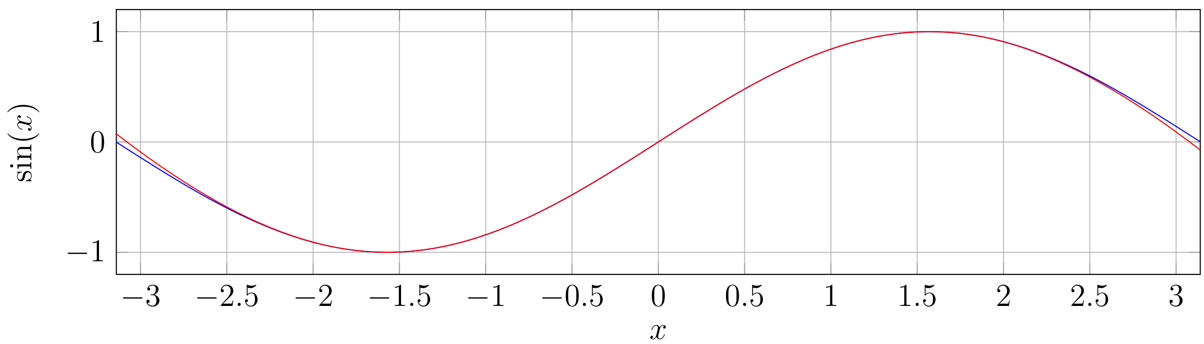
This is of course defeating the purpose. The real question is how to teach the computer what a sin function really is. Computers understand how to do couple primitive operations such as additon, subtraction, multiplication, division. If you combined these operations, you can compute any polynomial you want. For example, you can easily tell computer to calculate

$$f(x) = x^2 + 2x + 1$$

Let us plot a magical polynomail:

$$g(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}.$$

Let us also plot this polynomail against the sin function.



The two lines are almost exactly right on top of each other. This means that we can approximate the sin function at least for $x \in [-\pi, \pi]$ using $g(x)$ to a very good accuracy.

The coefficient of g indeed are not randomly picked. It's from Taylor's series of sin function.

Taylor Series

Let us consider a nice function[1] $f(x)$ for example $f(x) = \sin(x)$. If I were to claim that $f(x)$ can be written as a sum of infinite order polynomail.

$$f(x) = c_0 + c_1 x^1 + c_2 x^2 + c_3 x^3 + c_4 x^4 + \dots$$

Our job is to find $c_0, c_1, c_2, \dots$ in terms of f. First, c_0 is quite easy to find since if you plugin $x = 0$, you will be left with $f(0)$ on the left hand side and $only c_0$ on the right hand side.

$$f(0) = c_0 + c_1 \times 0 + c_2 \times 0^2 \dots = c_0$$

Thus, we can conclude that

$$c_0 = f(0)$$

One down infinitely more to go. Gettin c_1 is a bit tricky unless someone notice what happend when you diffferential both sides.

$$f'(x) = 0 + c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + \dots$$

With this we can just plug in $x = 0$ on both side and all the terms except c_1 are gone. Specifically,

$$f'(0) = 0 + c_1 + 0 + 0 + 0 + \dots$$

Therefore, we can conclude that

$$c_1 = f'(0)$$

Getting c_2 requires the same trick as c_1 all we need to do is differentiate it twice. For this we will use $f^{(i)}$ to denotes the i-th derivative of the function f.

$$f^{(2)}(x) = 0 + 0 + 2c_2 + 3 \times 2c_3 x + 4 \times 3c_4 x^2 + \dots$$

Then we plug in $x = 0$ to eliminate all the higher order terms. In the end we will be left with

$$f^{(2)}(0) = 2c_2 + 0 + 0 + \dots$$

Thus,

$$c_2 = \frac{f^{(2)}(0)}{2}$$

We can repeat the same exercise of differntialting and pluggin in 0 and conclude that

$$c_i = \frac{f^{(i)}(0)}{i!}$$

[1]Mathematically speaking you need a continuous and infinitely differentiable function in the domain we care about. Let us just focus on the core idea instead of mathematical details. If you are interested in the details look up the book or wikipedia.

So, any nice function f can be written as

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \dots \tag{1}$$
$$= f(0) + f^{(1)}(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \dots \tag{2}$$

This result is called *Taylor's Series.*

Let us consider the sin function shown above. The coefficient are the following:

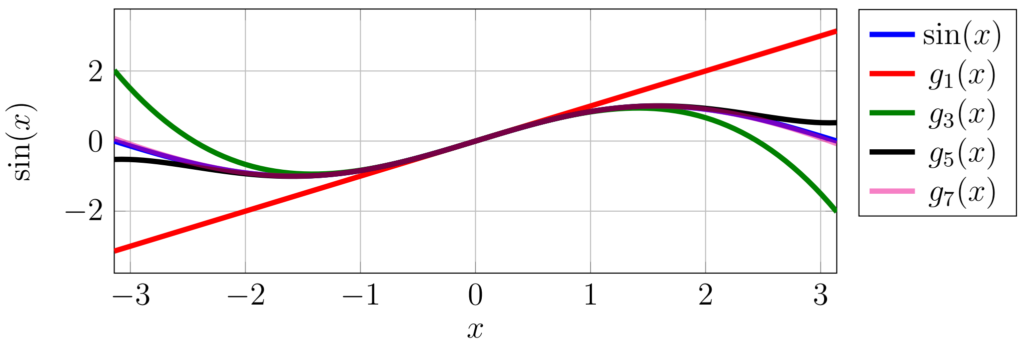
$$c_0 = \sin(0) = 0 \tag{3}$$
$$c_1 = \cos(0) = 1 \tag{4}$$
$$c_2 = -\sin(0)/2! = 0 \tag{5}$$
$$c_3 = -\cos(0)/3! = -1/6 \tag{6}$$
$$c_4 = \sin(0)/4! = 0 \tag{7}$$
$$c_5 = \cos(0)/5! = 1/120 \tag{8}$$
$$c_6 = -\sin(0)/6! = 0 \tag{9}$$
$$c_7 = -\cos(0)/7! = -1/5040 \tag{10}$$
$$\vdots$$

Thus, we can write

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

At this point you may scream: AJ., computers can't calculate infinite series. Yes, you are correct; it can not calculate infinite series. But, we can "hope" that if we truncate the series at some point we will not be too far away from the answer. Let us consider three polynomials where we just take the first few terms of the taylor series

$$g_1(x) = x \tag{11}$$
$$g_3(x) = x - \frac{x^3}{3!} \tag{12}$$
$$g_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} \tag{13}$$
$$g_7(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \tag{14}$$
$$\tag{15}$$



As you can see from the above plot, the more terms we added the closer we are when we are faraway from x. This is actually not a coincidence. We will quantify this later. One way to see this is because the additional terms in the series require higher power on x. If x is small, then x^2 is super small compared to x. So the additional terms you add it won't affect the value much.

Let us consider a more useful form of taylor series. Let us consider when x close to some fixed point x_0 but shifted by a little amount, h. Specifically,

$$x = x_0 + h \to x(h) = x_0 + h$$

Thus, we can parametize the curve in terms of shift amount h, instead of x.

$$f(x(h)) = f(x_0 + h) = g(h)$$

Then the Taylor series of $g(h)$ is given by

$$f(x_0 + h) = g(h) = g(0) + g^{(1)}(0)h + \frac{g^{(2)}(0)}{2!}h^2 + \frac{g^{(3)}(0)}{3!}h^3 + \frac{g^{(4)}(0)}{4!}h^4 + \dots$$

All the derivate of g evaluated at 0 is just the derivate of f evaluted at x_0. We can also obtain this result from chain rule

$$g(h) = f(x(h)) \to g'(t) = \frac{d}{dt}g(t) = \frac{d}{dt}f(x(t)) = f'(x(t))\frac{d}{dt}x(t) = f'(x(t))$$

We can do this to all the derivatives and conclude that

$$g^{(i)}(h) = f^{(i)}(x(h)) = f^{(i)}(x_0 + h)$$

Thus,

$$g^{(i)}(0) = f^{(i)}(x_0)$$

So the above series can be written as

$$f(x_0 + h) = f(x_0) + f^{(1)}(x_0)h + \frac{f^{(2)}(x_0)}{2!}h^2 + \frac{f^{(3)}(x_0)}{3!}h^3 + \frac{f^{(4)}(x_0)}{4!}h^4 + \dots$$

This is also called *Taylor's expansion of f around $x = x_0$*

Let us do on example:

$$f(x) = \ln(x)$$

If you try to expand this function around $x = 0$, you will be in trouble as $\ln(0)$ is undefined. The proper thing to do is to expand it around $x = 1$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

This is one of the most useful expansion out there. It is needed for log likelihood minimization. Many languages have this function implemented for you; for example, in Python you can use `math.log1p`.

Taylor Theorem

For this section don't worry too much about the prove, I need you to understand the result.

One of the most important question one can ask is that if we truncate the series after then n-th term how accurate is our estimate. We know that

$$f_3(x) = x - \frac{x^3}{3}$$

and

$$f(x) = \sin(x)$$

are really close. They only deviate when x is large enough. So, it would be nice to bound the difference

$$|f_3(a) - f(a)| \leq R(a).$$

This information would tell us how many terms we need to achieve a given accuracy.

To calculate the bound, let us apply the first fundamental theorem of calculus successively. Let use first remind you what the first fundamental theorem of calculus is

$$\int_{t_1=a}^{x} f'(t_1)\, dt_1 = f(x) - f(a)$$

Thus, we have

$$f(x) = f(a) + \int_{t_1=a}^{x} f'(t_1)\, dt_1. \tag{16}$$

Let us do it again. This time we will apply the first fundamental theorem of calculus on $f'(t)$.

$$f'(t_1) = f'(a) + \int_{t_2=a}^{t_1} f''(t_2)\, dt_2 \tag{17}$$

You should read this line carefully twice and make sure you understand all the limits and all the primes.

Plugging 17 in to 16 we have

$$\begin{aligned}
f(x) =& f(a) + \int_{t_1=a}^{x} \left[f'(a) + \int_{t_2=a}^{t_1} f''(t_2)\, dt_2 \right] dt_1 && (18)\\
=& f(a) + \int_{t_1=a}^{x} f'(a)\, dt_1 + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} f''(t_2)\, dt_2\, dt_1 && (19)\\
& && (20)
\end{aligned}$$

The $f'(a)$ in second term is just a constant for that integral[2]. We are now left with

$$\begin{aligned}
f(x) =& f(a) + f'(a)(x-a) + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} f''(t_2)\, dt_2\, dt_1 && (21)\\
& && (22)
\end{aligned}$$

[2] $\int_{t_1=a}^{x} f'(a)\, dt_1 = f'(a) \int_{t_1=a}^{x} dt_1 = f'(a)(x-a)$

We can then do the same trick again for $f''(t_2)$

$$f''(t_2) = f''(a) + \int_{t_3=a}^{t_2} f^{(3)}(x)\, dt_3$$

Plugging this back in to 21 yields

$$\begin{aligned} f(x) =& f(a) + f'(a)(x-a) + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \left[f''(a) + \int_{t_3=a}^{t_2} f^{(3)}(x)\, dt_3 \right] dt_2\, dt_1 && (23)\\ =& f(a) + f'(a)(x-a) + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} f''(a)\, dt_2\, dt_1 + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \int_{t_3=a}^{t_2} f^{(3)}(x)\, dt_3\, dt_2\, dt_1 && (24) \end{aligned}$$

For the third term, again, $f''(a)$ is just a constant so we have

$$\begin{aligned} \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} f''(a)\, dt_2\, dt_1 =& f''(a) \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} dt_2\, dt_1 && (25)\\ =& f''(a) \int_{t_1=a}^{x} (t_1 - a)\, dt_1 && (26)\\ =& f''(a) \int_{u=0}^{x-a} u\, du \qquad \text{where } u = t_1 - a && (27)\\ =& f''(a) \frac{u^2}{2} \Big|_{u=0}^{u=x-a} && (28)\\ =& f''(a) \frac{(x-a)^2}{2} && (29) \end{aligned}$$

So, we have

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2} + \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \int_{t_3=a}^{t_2} f^{(3)}(x)\, dt_3\, dt_2\, dt_1 \quad (30)$$

We can keep repeating applying the first fundamental theorem of calculus. By now you should be convinced that we will be left with something like

$$\begin{aligned} f(x) =& f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2} + \ldots + f^{(n)}(a)\frac{(x-a)^n}{n!} + \\ & \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \cdots \int_{t_{n+1}=a}^{t_n} f^{(n+1)}(x)\, dt_{n+1} \ldots dt_2\, dt_1 && (31) \end{aligned}$$

This result is called *Taylor's Theorem*. One of the most important theorem in numerical method. We will come back to this formula again and again.

We call the first part of 31 the *Polynomal term*

$$P_n(x) \equiv f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2} + \ldots + f^{(n)}(a)\frac{(x-a)^n}{n!}. \quad (32)$$

This is exactly the truncated Taylor Series.

The latter term of 31 is called the *Remainder term.*

$$R_n(x) \equiv \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \cdots \int_{t_{n+1}=a}^{t_n} f^{(n+1)}(x)\, dt_{n+1} \ldots\, dt_2\; dt_1 \tag{33}$$

Bounding the Remainder

This remainder expression looks super scary. But it is a very important term as it encapsulates the difference between the true function and the truncated series. Specifically,

$$f(x) = P_n(x) + R_n(x) \tag{34}$$

Our real aim is to bound the scary looking expression R_n. So let us do that. The idea is that if the $f^{(n+1)}(x)$ within the regiod of $[a, x]$ is bounded then we use that to the bound the multiple integral int he remainder. So let us assume that for $t_{n+1} \in [a, x]$

$$m \leq f^{n+1}(t_{n+1}) \leq M$$

Let us apply the scary looking integral on all the terms[3]

$$\int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \cdots \int_{t_{n+1}=a}^{t_n} f^{(n+1)}(x)\, dt_{n+1} \ldots\, dt_2\; dt_1 \leq \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \cdots \int_{t_{n+1}=a}^{t_n} M\, dt_{n+1} \ldots\, dt_2\; dt_1 \tag{35}$$

$$\leq M \frac{(x-a)^{n+1}}{(n+1)!} \tag{36}$$

The left hand side give you the same thing so we have[4]

$$m\frac{(x-a)^{n+1}}{(n+1)!} \leq \int_{t_1=a}^{x} \int_{t_2=a}^{t_1} \cdots \int_{t_{n+1}=a}^{t_n} f^{(n+1)}(x)\, dt_{n+1} \ldots\, dt_2\; dt_1 \leq M\frac{(x-a)^{n+1}}{(n+1)!} \tag{37}$$

This means that the remainder is bounded by a simple formula

$$m\frac{(x-a)^{n+1}}{(n+1)!} \leq R_n(x) \leq M\frac{(x-a)^{n+1}}{(n+1)!} \tag{38}$$

This form is quite inconvenient to use. Let us workout a bit more to get it in to a more useful form. The left hand side and the right hand side just differ by the number multiplying in front. This means, that

$$R_n(x) = K\frac{(x-a)^{n+1}}{(n+1)!} \;\; \exists K \in [m, M]$$

[3] We will only do the right hand side for brevity. The left hand side looks exactly the same.

[4] One actually need to be careful a little bit to swith the sign when $n+1$ is odd and $x-a$ is negative.

Here is the magic part, since m is the minimum of $f^{(n+1)}(x)$ for $x \in [a, x]$ and M is maximum of $f^{(n+1)}(x)$ for $x \in [a, x]$. By continuity of $f^{(n+1)}(x)$, there is a point $\xi \in [a, x]$ such that

$$f^{(n+1)}(\xi) = K$$

This allows us to write the remainder term as

$$R_n(x) = f^{(n+1)}(\xi)\frac{(x-a)^{n+1}}{(n+1)!} \quad \exists \xi \in [a, x]$$

Combining this with taylor series we got earlier we have the most important formula in this class, the Taylor Theorem.

$$\boxed{f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2} + \ldots + f^{(n)}(a)\frac{(x-a)^n}{n!} + f^{(n+1)}(\xi)\frac{(x-a)^{n+1}}{(n+1)!}}$$

for some $\xi \in [a, x]$.

Let us use it

The error bound on truncated taylor series can be found using

$$f(x) = P_n(x) + R(x)$$

where

$$\left|f^{(n+1)}(t)\right| \leq M \text{ for } t \in [x, a] \tag{39}$$

Then,

$$|R_n(x)| \leq M\frac{(x-a)^{n+1}}{(n+1)!} \tag{40}$$

Let us do an example to appreciate this. Suppose that we are trying to calculate

$$f(x) = \ln(x)$$

at $x = 1.01$ using the tyalor expansion of $\ln(x)$ around $x = 1$.

$$f_2(x) = (x-1) - \frac{(x-1)^2}{2}$$

We are intrested in how far off we will be from the real answer. The real difference can be calculated just by

$$\text{real diff} = \ln(1.01) - f_2(0.01)$$

But this is NOT what we are looking for. Since the computer has no idea about the real ln function, the best we can say is the bound on the remainder.

$$\ln(x) = f_2(x) + R_2(x)$$

So the bound on remainder we need to calculate is

$$|R_2(x)| \leq M\frac{(x-1)^3}{3!}$$

where

$$\left|f^{(3)}(t)\right| \leq M \text{ where } t \in [1, 1.01]$$

Three things to note

1. The it's the real ln function that we do triple differentiation on and find the bound and NOT the approximate function.
2. Second, it's $(x-1)^3$ because our approximate function is expanded around $x = 1$.
3. Third, the range where we find the maximum is $t \in [1, 1.01]$ since we use expand it around 1 and we want to find the value at 1.01.

The third derivative of ln function is

$$\frac{d^3}{dt^3} \ln(x) = \frac{2}{t^3}$$

The maximum of this occurs when $t = 1$. So

$$M = \frac{2}{1^3} = 2$$

This means that the remainder is

$$R_2(1.01) \leq 2 \times \frac{0.01^3}{3!} \leq 3.33 \times 10^{-7}$$

That we can say the following about the approximation for $\ln(1.01)$ with $(x-1)-(x-1)^2/2$

$$\ln(1.01) \approx 0.01 - \frac{0.01^2}{2} \approx 0.009\,950 \pm 0.000\,000\,333\ldots$$

The true value of $\ln(1.01)$ is

$$\ln(1.01) = 0.009\,950\,330\,853\,168\ldots$$

This is actually quite amazing since we can calculate $\ln(1.01)$ to correct 6 digit with only two terms.

You can see the the major factor in making the error small is the fact that we have 0.01^3 term in the error bound. We normally write the formula as

$$\ln(1+h) = h - \frac{h^2}{2} + \mathcal{O}(h^3)$$

Another Example

The real use of the bound on the remainder is in this example. Let us look at the previous problem the other way around.This time we are be given a needed accuracy and we need to figure out how many terms we need. For example, suppose we want to calculate $\sin(2)$ that is accurate up to 10^{-7}, how many terms of the taylor series do we need.

First, recall that

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

If we stop the series at x^7 we actually get x^8 terms for free since it's 0. So,

$$P_7(x) = P_8(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}$$

This means that instead of using R_7 as error we can just use R_8 as the bound on the error for the above approximation.

Since the nth derivate just cycle between sine and cosine. Plus sine and cosine is bounded by 1 and -1. The n-th remainder is givenby

$$|R_n(x)| \leq 1 \times \frac{x^{n+1}}{n+1}$$

Our goal is to find n such that $|R_n(2)| \leq 10^{-7}$. This one is easy we can just us brute for this one out.

	Bound
R_2	$\frac{2^3}{3!} = 1.33$
R_4	$\vdots$
R_6	$\vdots$
R_8	$\vdots$
R_{10}	$\frac{2^{11}}{11!} = 5 \times 10^{-5}$
R_{12}	$\vdots$
R_{14}	$\frac{2^{15}}{15!} = 2.5 \times 10^{-8}$

So the approximation that will give us $\sin(2)$ at the accuracy of at least 10^{-7} is

$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \frac{x^{13}}{13!} + \mathcal{O}(x^{15})$$

which is

$$\sin(x) \approx 0.909\,297\,451\,519\,674$$

The true value is

$$\sin(x) = 0.909\,297\,426\,825\,682$$

That's actually a lot of terms and the main reason is because the point where we want to caculate is so far off from the expansion point.

Numerical Precision

You may ask why do we need an extra function to accurately calcuate $\ln 1 + 1e - 50$ why don't we just use ln function instead of calling a complicated $ln1p(1e-7)$ a.

The reason lies in how computer store floating points. Without going into too much details [5], the computer store floating point as sign, mantissa and exponent and it has limited number of bit reserved for each part. So, a number like $-0.000\,000\,000\,110\,111\,101_2$

[5]See `https://en.wikipedia.org/wiki/Single-precision_floating-point_format`

will be splitted into three parts

$$\underbrace{-1}_{sign} \times 1.\underbrace{10111101}_{fraction} \times 2^{\overbrace{exponent}^{-10}}$$

That is to represent a floating point you need to store three number: -1 for sign, 10111101 for the fraction, and -10 for the exponent.

For IEEE754 single precision dictates that 1bit is for sign, 8 bit will be used for exponent and 23 bit will be used for fraction. This means that there are a lot of number that can't be represented by floating point. For example, you can't represent `1+1e-50` with single precision number because it requires 50 bit for the fraction part. This also means you can do funny thing like

```
(1.+1e-50) - 1.
```

This will give you zero and not the mathematically correct 1e-50. This is because python try to calculate `1+1e-50` and put it in a floting format which it can't. So, the number got truncated and stored in the ram as `1.`. Then we subtract it from it 1, we got zero as the result.

As an exercise try to explain why this returns false in python

```
1.+1e-50-1. == 1.-1.+1e-50
```

There is a whole literature teaching you how to count the loss of precision and such. I just want you to be aware of such issue. If you run into some funny results, this may be the reason. The rule of thumb is that try to do operation of the number that is of the same order of magnitude first.

Finding Constants

Another use for Taylor Series is for calculating constants. This is not exactly that way people do it but it doesn't hurt to know. We can express the value for the constant π in terms of infinte series. This can be done by noticing that

$$\tan\frac{\pi}{4} = 1 \rightarrow \pi = 4 \times \arctan 1$$

So all we need to do is calculate the taylor series of arctan around 0. You can look up the table and find that [6]

[6] Look up wikipedia how differentiation of arctan comes about. If you forgot about implicit rule, that is.

$$
\begin{aligned}
\frac{d}{dx}\arctan(x) &= \frac{1}{1+x^2}\\
\frac{d^2}{dx^2}\arctan(x) &= \frac{1}{(1+x^2)^2}\times 2x\\
\frac{d^3}{dx^3}\arctan(x) &= \frac{6x^2-2}{(1+x^2)^3}\\
\frac{d^4}{dx^4}\arctan(x) &= \frac{24x(x^2-1)}{(1+x^2)^4}\\
\frac{d^5}{dx^5}\arctan(x) &= \frac{24(5x^4-10x^2+1)}{(1+x^2)^5}\\
\frac{d^6}{dx^6}\arctan(x) &= \dots\\
&\vdots
\end{aligned}
$$

205 Thus the taylor series of arctan around $x = 0$ is given by[7]

$$\arctan(x) = 0 + x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots$$

206 This implies

$$\pi = 4\arctan(1) = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots\right)$$

[7]There is actually a bit smarter way to do this. See `https://www.math.hmc.edu/funfacts/ffiles/30001.1-3.shtml`