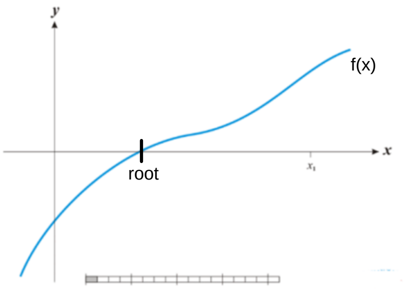
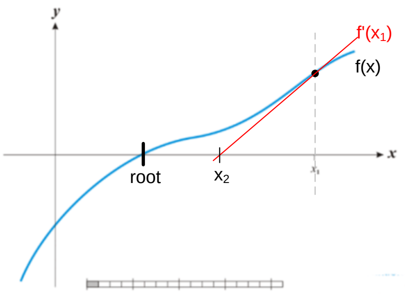
Newton’s method

參考網站：
https://ccjou.wordpress.com/2013/07/08/牛頓法——非線性方程的求根方法/

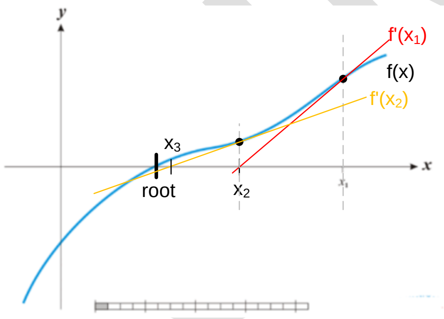
為尋找解的一種方法，即找尋 $f(x)=0$時的x



上課解釋：



$f'(x_1)=\dfrac{f(x_1)}{x_1-x_2}$，$f'(x_1)$、$f(x_1)$、$x_1$都是已知，可推得$x_2=x_1-\dfrac{f(x_1)}{f'(x_1)}$
以此類推，

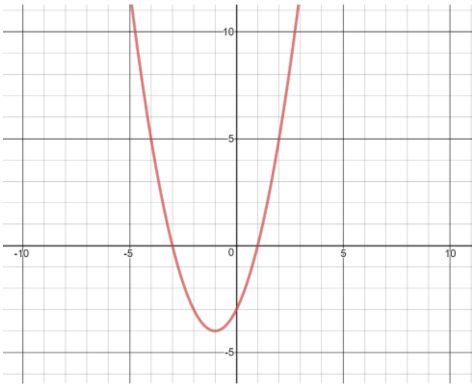


$f'(x_2)=\dfrac{f(x_2)}{x_2-x_3}$，$f'(x_2)$、$f(x_2)$、$x_2$都是已知，可推得$x_3$=$x_2-\dfrac{f(x_2)}{f'(x_2)}$

一直重複同樣的事情直到 $f'(x_i)=0$ 或是x_i-x_{i-1}在可容許的誤差內時(e.g.小數第三位都相同)

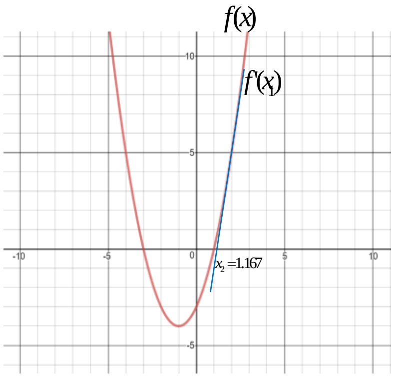
e.g.
$f(x)=x^2+2x-3$的圖形如下，我們可預先知道解為 1 和-2，來看牛頓法求出的解為何，假設我們能

容許誤差為小數第二位

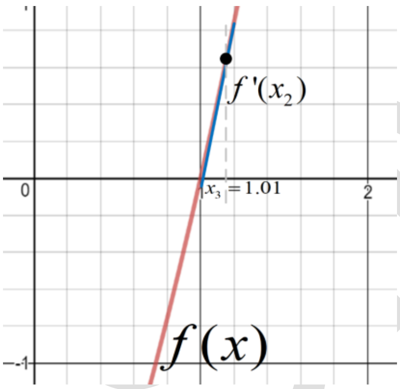


假設我們隨便取 $x_1=2$

$$x_2=x_1-\frac{f(x_1)}{f'(x_1)}=2-\frac{5}{6}=1.167$$



再迭代一次， $x_3=x_2-\dfrac{f(x_2)}{f'(x_1)}=1.167-\dfrac{0.696}{4.334}=1.01$



以下的迭代就不做了，整理一個表格出來

i	x_i	f(x)	f'(x)	x_{i+1}
2	1.01	0.04	4.02	1.00
3	1.00	0.00	4.00	1.00
4	1.00	0.00	4.00	1.00

可看到最終我們可收斂至解為 1

網站解釋(和上課有稍許不同而已)
以泰勒級數顯示 f(x)

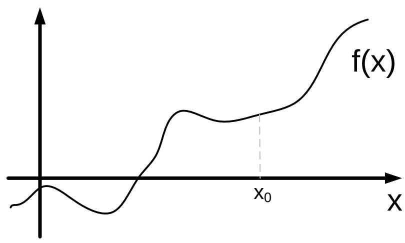
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + O((x - x_0)^2)$$

我們欲解 $f(x) = 0 = f(x_0) + f'(x_0)(x - x_0) + O((x - x_0)^2)$

我們忽略 2 次方以上的項，也就是將 $f(x)$ 近似並退化成一個一次多項式，此時就變成解下列問題：
$f(x_0) + f'(x_0)(x - x_0) = 0$

$x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$ ，和剛剛推導出的式子是一樣的

Taylor series



$f(x)$在x_0做展開

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f'''(x_0)(x - x_0)^3 + \ldots$$
$$= \sum_n \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

只做說明不做證明：
$f(x_0) = f(x_0) + 0 + 0 + 0 + \ldots = f(x_0)$
$f'(x_0) = 0 + f'(x_0) + 0 + 0 + \ldots = f'(x_0)$
$f''(x_0) = 0 + 0 + f''(x_0) + 0 + \ldots = f''(x_0)$
...
在x_0這點而言，這個方程式無論微分多少次都成立

若$x_0 = 0$, 又稱為Maclaurin series

e.g.
一些常用的級數

$$(x_0{=}0)\ \ e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \ldots = \sum_n \frac{x^n}{n!}$$

$$(x_0=0)\ \sin x = \sin 0 + x\cos 0 + \frac{1}{2}x^2(-\sin 0) + \frac{1}{3!}x^3(-\cos 0) + \ldots = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \ldots$$

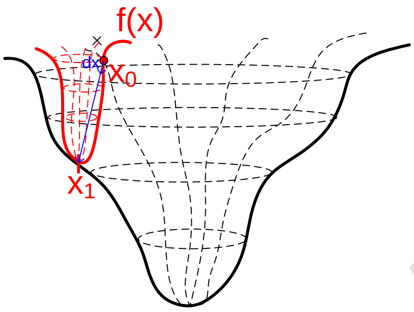
$$(x_0=0)\ \cos x = \cos 0 + x\sin 0 + \frac{1}{2}x^2(-\cos 0) + \frac{1}{3!}x^3(-\sin 0) + \ldots = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \ldots$$

$$\Rightarrow e^{ix} = 1 + ix + \frac{1}{2}(-1)x^2 + \frac{1}{3!}(-i)x^3 + \ldots$$

$$= (1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \ldots) + i(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \ldots) = \cos x + i\sin x$$

Newton’s Method in optimization

有了 Newton’s Method，也有了 Taylor series，我們可以利用 Newton’s Method 來做 optimization
https://ch-hsieh.blogspot.tw/2012/04/generalized-newton-raphson-algorithm.html
https://zh.wikipedia.org/wiki/應用於最優化的牛頓法



代數意義

$$f(x_0) = f(x_0+dx) = f(x_0) + f'(x_0+dx)(x_0+dx-x_0) + \frac{1}{2!}f''(x_0)(x_0+dx-x_0)^2 + \ldots$$

$$\approx f(x_0) + f'(x_0)dx + \frac{1}{2}f''(x_0)(dx)^2$$

剛剛我們是找 $f(x)=0$時的x，這裡我們要找的是 $f'(x)=0$ 的x

一樣從 x_0 出發，我們希望到找到 $f'(x_0+dx)=0$的 x，即圖中的點 x_1，我們將等式左右兩端微分後=0 得

$$\frac{df(x_0+dx)}{d(dx)} = f'(x_0) + f''(x_0)dx = 0$$

$$\Rightarrow dx = \frac{-f'(x_0)}{f''(x_0)}$$

$$x_{new} = x_0 + dx = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

若空間維度不只一維

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = x_0 - Hf(x_0)^{-1} \nabla f(x_0)$$

其中 $Hession\ \ function\ of\ \ f(x) = Hf(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_0}^2} & \frac{\partial^2 f}{\partial x_0 \partial x_1} & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_0} & \frac{\partial^2 f}{\partial {x_1}^2} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$

我們不斷的重複上述動作，直至 $x_i - x_{i-1}(dx)$的值小至我們可忽略時方可結束

幾何意義

應該蠻多人無法將上式的 $x_{new} = x_0 + dx = x_0 - \frac{f'(x_0)}{f''(x_0)}$ 和上面的圖聯想在一起，不過，我們可以先看這個式子

$$f(x_0) = f(x_0 + dx) \approx f(x_0) + f'(x_0)dx + \frac{1}{2} f''(x_0)(dx)^2$$

這是一個二次方程式，我們是將經過 x_0 的多項式 $f(x)$ 簡化為二次多項式，也就是上圖中紅色的

$f(x)$。然後，我們再求這個二次多項式的極值時的 x，這個 x 就是 x_1，這個方法不是很神奇的馬上就找到在 $f(x)$ 上最靠近自己的微分=0 的點，是找經過自己的近似二次多項式的微分=0 的點，有點繞口，希望你們能懂。

驗證(LSE)

我們得到一個得到極大解或極小解的近似方法，我們就用在前一堂課學到的 LSE 上試試看，由於 LSE 是一個二次多項式，所以 Taylor series 展開的最高次方只會到兩次，其近似的二次多項式就是實際的 LSE 多項式，解出來微分=0 的點就是實際 LSE 的最小值，故做一次就應該到最小值了。

$$LSE = \| A\vec{x} - \vec{b} \|^2 = x^T A^T A\vec{x} - 2x^T A^T \vec{b} + \vec{b}^T \vec{b} = f(\vec{x})$$

$$\nabla f(\vec{x}) = 2A^T A\vec{x} - 2A^T \vec{b}$$

$$Hf(\vec{x}) = 2A^T A$$

$$\begin{aligned} \overline{x_1} &= \overline{x_0} - (2A^T A)^{-1}(2A^T A\overline{x_0} - 2A^T \vec{b}) \\ &= \overline{x_0} - (A^T A)^{-1} A^T A\overline{x_0} + (A^T A)^{-1} A^T \vec{b} = (A^T A)^{-1} A^T \vec{b} \end{aligned}$$

其實根據上一堂課的結論，做到這一部已經是 global 的最小解了，為了證明已經收斂，我們再做一次 Newton method

我們可以看到上式，無論 $\vec{x}$ 為何，該 step 的最小值都會是 $(A^T A)^{-1} A^T \vec{b}$，故如果我們將第一次 step 找到的最小值當作第二次 step 的初始值，第二次 step 得到的最小值仍會是 $(A^T A)^{-1} A^T \vec{b}$

$$LSE = \| A\overline{x_1} - \vec{b} \|^2 = \overline{x_1}^T A^T A\overline{x_1} - 2\overline{x_1}^T A^T \vec{b} + \vec{b}^T \vec{b} = f(\vec{x})$$

$$\nabla f(\overline{x_1}) = 2A^T A\overline{x_1} - 2A^T \vec{b}$$

$$Hf(\overline{x_1}) = 2A^T A$$

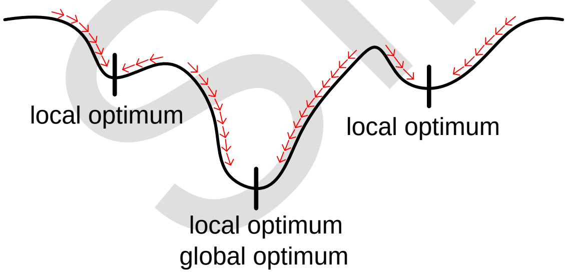
$$\begin{aligned} \overline{x_2} &= \overline{x_1} - (2A^T A)^{-1}(2A^T A\overline{x_1} - 2A^T \vec{b}) \\ &= \overline{x_1} - (A^T A)^{-1} A^T A\overline{x_1} + (A^T A)^{-1} A^T \vec{b} = (A^T A)^{-1} A^T \vec{b} \end{aligned}$$

故只要做一次 Newton Method 就收斂

此方法缺點：

1. 若空間維度太高，Hession function 的維度會太高(若有 n 維，會產生 nxn 維的矩陣)，做 Hession function 反矩陣會耗費 $O(n^3)$的時間
2. 只能找到 local optimization，若維度太高，容易陷入局部的極小值(local optimum)

e.g.



紅色箭頭是各點的 dx 方向，一旦到 local optimum 後，就不可能離開該點了
但是相較於其他的近似法，Newton Method 可使用最少次的迭代即可收斂

note:
若 Hession function 不可逆，可使用 pseudo inverse