2018801013
Subhajit Saho



# HW5

① prove the following using jensen's inequality

$$\frac{1}{x-1} + \frac{1}{x} + \frac{1}{x+1} > \frac{3}{x}$$

use this to show that

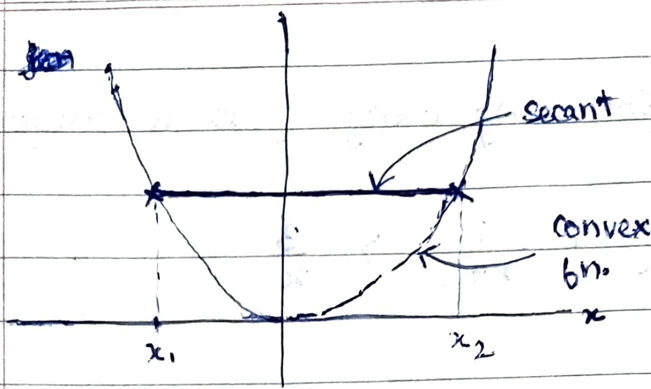
$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

does not converge to a real no.

A) i wanted do reality check first. this is a javascript fn. to have some values.

```
function reciprocalSum (n) {
    return n>0? reciprocalSum(n-1) + 1/n
        : 0;
}
```

| | | | | |
|---|---|---|---|---|
| reciprocalSum (1) | | = | 1 | not converging |
| ... | 10 | = | 2.9 | |
| ... | 100 | = | 7.5 | |
| ... | 1000 | = | 9.8 | |




as we can see from above, between 2 points $x_1, x_2$ the secant is always above the convex fn.

i.e., for any point $x$ between $x_1$ and $x_2$

$$x = a x_1 + (1-a) x_2$$

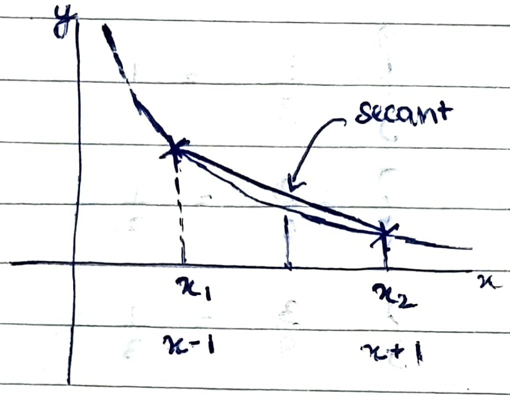
we can say that,

$$f_{secant}(x) \geq f_{convex}(x)$$

or, $a f(x_1) + (1-a) f(x_2) \geq f(a x_1 + (1-a) x_2)$

this is jensen's inequality.

now to look at $f(x) = \frac{1}{x}$




also convex fn.

also secant $\geq$ fn.

for the midpoint between $x_1 = x-1$
$x_2 = x+1$

$$\frac{1}{2}\left(\frac{1}{x-1}\right) + \frac{1}{2}\left(\frac{1}{x+1}\right) \geq \frac{1}{x} \quad f(x)$$

$$\Rightarrow \frac{1}{x-1} + \frac{1}{x+1} \geq \frac{2}{x}$$

$$\Rightarrow \frac{1}{x-1} + \frac{1}{x} + \frac{1}{x+1} \geq \frac{3}{x}$$

Since we know $f(x) = \frac{1}{x}$ is not a straight line,

$$\frac{1}{x-1} + \frac{1}{x} + \frac{1}{x+1} > \frac{3}{x} \qquad \text{(proved)}$$

$1$

$\left. + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \right\} \quad \frac{3}{3} = 1$

$\left. + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right\} \quad \frac{3}{6} = \frac{1}{2}$

$\left. + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} \right\} \quad \frac{3}{9} = \frac{1}{3}$

$\left. + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} \right\} \quad \frac{3}{12} = \frac{1}{4}$

$+ \ldots$

So, if $S = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$

we see that $S = 1 + 3 \times S$

- Since $S$ is positive, we can say that it is <u>not convergent</u>.

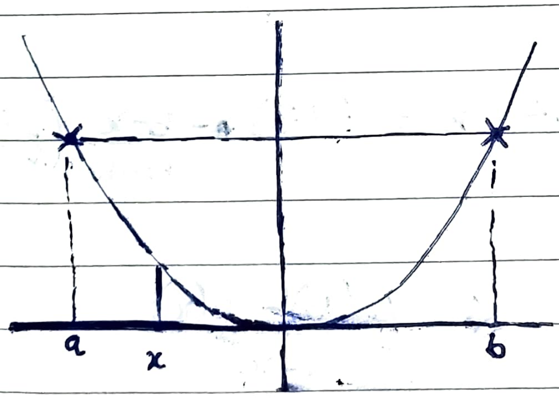
② prove the 3-chord lemma

if $f:[a,b] \to \mathbb{R}$ is convex, and

$a < x < b$, then

$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(a)}{b-x}$$

first, lets do a reality check




here, $\frac{f(x)-f(a)}{x-a}$ is -ve

$\frac{f(b)-f(a)}{b-a}$ is 0

$\frac{f(b)-f(x)}{b-x}$ is +ve

satisfies..

using the rule,

secant $\geq$ convex fn. (at $x$)

$$f(a) + \frac{x-a}{b-a}\left(f(b) - f(a)\right) \geq f(x)$$

$$\Rightarrow \frac{x-a}{b-a}\left(f(b) - f(a)\right) \geq f(x) - f(a)$$

$$\Rightarrow \frac{f(b) - f(a)}{b-a} \geq \frac{f(x) - f(a)}{x-a} \quad \text{--- (1)}$$

also, secant at $x = f(b) - \frac{b-x}{b-a}\left(f(b) - f(a)\right)$

$$\Rightarrow \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x} \quad \text{--- (2)}$$

so we can say,

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x}$$

(3) if $x, y, z$ are +ve real no. such that

$x + y + z = 1$,

show using jensen's inequality that

$$\left(1+\frac{1}{x}\right)\left(1+\frac{1}{y}\right)\left(1+\frac{1}{z}\right) \geq 64.$$

attempt1:

$$\left(\frac{2x+y+z}{x}\right)\left(\frac{x+2y+z}{y}\right)\left(\frac{x+y+2z}{z}\right) \cancel{\geq 6}$$

$$= \left(2+\frac{y+z}{x}\right)\left(2+\frac{x+z}{y}\right)\left(2+\frac{x+y}{z}\right)$$

$\geq 8$

(but not good enough)

attempt2: (with help)

let $f(x) = \log\left(1+\frac{1}{x}\right)$ (convex fn.)

using the role

Secant $\geq$ convex fn. at K

where $K = \frac{1}{3}(x+y+z)$

$$\frac{1}{3}\{f(x)+f(y)+f(z)\} \geq f\left(\frac{x+y+z}{3}\right)$$

$$\frac{1}{3}\log\left(1+\frac{1}{x}\right)\left(1+\frac{1}{y}\right)\left(1+\frac{1}{z}\right) \geq f(1/3)$$

$$\log\left(1+1/x\right)\left(1+1/y\right)\left(1+1/z\right) \geq 3\log\left(1+\frac{1}{1/3}\right)$$

$$= 3\log 4$$

$$= \log 64$$

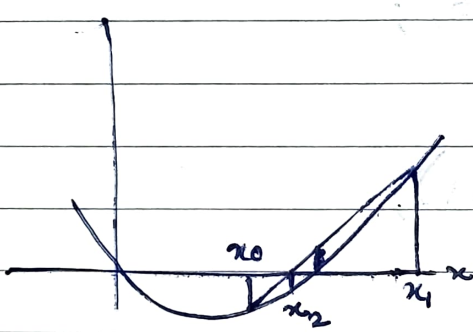
$$\Rightarrow \left(1+\frac{1}{x}\right)\left(1+\frac{1}{y}\right)\left(1+\frac{1}{z}\right) \geq 64$$

(4) conduct 2 iterations of the secant method

$$x^3 - 2x - 5 = 0$$

$x_0 = 2 \qquad x_1 = 3$

secant method is a root finding iterative algorithm. it was developed long before newton's method.




* it is better ~~...~~a version of bisection method.

$$x_2 = x_1 - f(x_1) \frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

derivation:

Solenoid

$$y_2 = 0$$

$$= \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x_2 - x_1) + f(x_1)$$

to make calculations simple, using javascript (nodejs).

```
f = x => x**3 - 2*x - 3;

x = (x1, x0) => x1 - f(x1) * ((x1-x0)/(f(x1)-f(x0)));

x0 = 2;
x1 = 3;

x2 = x(x1, x0);
```
↳ 1.94

```
x3 = x(x2, x1);
```
↳ 1.92

⑤ conduct 2 iterations of the bisection method for

$x e^{x} = 1$ where $x \in [0, 1]$

```
f = x => x * Math.exp(x) - 1;
```

f(0)
↳ -1

f(0.5)
↳ -0.18

f(1)
↳ 1.72

f(0.75)
↳ 0.59

f(0.625)
↳ 0.17

getting close

(6) find a real root of the equations

$x^2 - y^2 = 3$ — (1)

$x^2 + y^2 = 13$ — (2)

by doing 2 iterations of newton's method

$x_0 = y_0 = \sqrt{6.5}$

$f_1 = x^2 - y^2 - 3$

$\frac{\partial f_1}{\partial x} = 2x \qquad \frac{\partial f_1}{\partial y} = -2y$

$\therefore \quad x_1 = x_0 - \frac{f_1(x_0, y_0)}{2x_0}$

$y_1 = y_0 + \frac{f_1(x_0, y_0)}{2y_0}$

so no lets put into javascript:

$x_1 = 3.14$

$y_1 = 1.96$

$x_2 = 2.66$

$y_2 = 2.73$

$x_3 =$

$y_3 =$

$$f_2 = x^2 + y^2 - 13$$

$$\frac{\partial f_2}{\partial x} = 2x \qquad \frac{\partial f_2}{\partial y} = 2y$$

$$\therefore \quad x_1 = x_0 - \frac{f_2(x_0, y_0)}{2x_0}$$

$$y_1 = y_0 - \frac{f_2(x_0, y_0)}{2y_0}$$

so,

$$x_1 = 2.55$$
$$y_1 = 2.55$$

$$x_2 = 2.55$$
$$y_2 = 2.55$$

looks like already at root

7 do 2 iterations of the fixed point method with

$$2x - \cos x - 3 = 0$$

take $g(x) = \dfrac{\cos x + 3}{2}$ , $x_0 = \dfrac{\pi}{2}$

$$\left.\begin{aligned} x_1 &= 1.5 \qquad & f(x_1) &= -0.07 \\ x_2 &= 1.54 \qquad & f(x_2) &= 0.04 \end{aligned}\right\} \text{in nodejs}$$

its coming close to the root.

(8) write the pseudocode for the gradient ascent algorithm. your code must be vectorized. also run the algorithm to find the maximum of

$$6 - (x_1^2(x_1^2 - 16) + x_2^2(x_2^2 - 9)) \quad \text{from } (0,0)$$

in gradient descent,

$$x_1 = x_0 - \eta \nabla f(x_0)$$

$$\nabla f = \begin{bmatrix} -4x_1^3 + 32x_1 \\ \\ -4x_2^3 + 18x_2 \end{bmatrix}$$

```
function gradDes (gradf, x0, n=0.001) {
  do {
    var x1 = x0.map((v,i) => v-n*gradf[i](x0));
    var e = x1.reduce((acc,v,i) =>
                  acc + Math.abs(v-x0[i]),0);
    x0 = x1;
    console.log({x1,e});
  } while (e > 0.000001);
  return x1;
}
```

gradDes (gradf, x0) → [0,0]

} it looks like (0,0) could be maxima, as it appears for other values of $x_0$ as well.

classmate
Date 20.4.2020
Page 16

(9) many people prefer to use the gradient descent algorithm in batch mode. how and why do you think they do that?

most people actually use mini-batch gradient descent of several samples: it is computationally a whole lost faster, and does not require holding large datasets in RAM.

(10) give a proof of the convergence of the gradient descent algorithm. also, discuss how to choose and adjust the learning rate so that it converges.

assume that $f: \mathbb{R}^n \to \mathbb{R}$
is convex and differentiable, and

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \text{for any } x, y$$

i.e; $\nabla f$ is lipschitz continuous with constant $L > 0$

gradient descent with fixed step size $t \leq 1/L$ satisfies

$$f(x^{(k)}) - f(x^*) \leq \frac{\|x^{(0)} - x^*\|^2}{2tk}$$

i.e, gradient descent has convergence rate $O(1/k)$

i.e., to get $f(x^{(k)}) - f(x^*) \leq \epsilon$,

need $O(1/\epsilon)$ iterations.

$\nabla f$ lipschitz with constant $L \Rightarrow$

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2 \quad \forall_{x,y}$$

plugging in $y = x - t\nabla f(x)$

$$f(y) \leq f(x) - \left(1 - \frac{Lt}{2}\right) t \, \|\nabla f(x)\|^2$$

letting $x^+ = x - t\nabla f(x)$ and taking $0 < t \leq 1/L$

$$f(x^+) \leq f(x^*) + \nabla f(x)^T (x - x^*) - \frac{t}{2}\|\nabla f(x)\|^2$$

$$= f(x^*) + \frac{1}{2t}\left(\|x - x^*\|^2 - \|x^+ - x^*\|^2\right)$$

Summing over iterations

$$\sum_{i=1}^{K} \left(f(x^{(i)}) - f(x^*)\right)$$

$$\leq \frac{1}{2t}\left(\|x^{(0)} - x^*\|^2 - \|x^{(K)} - x^*\|^2\right)$$

$$\leq \frac{1}{2t}\|x^{(0)} - x^*\|^2$$

since $f(x^{(K)})$ is non-increasing

$$f(x^{(K)}) - f(x^*) \leq \frac{1}{K}\sum_{i=1}^{K}\left(f(x^{(i)}) - f(x^*)\right)$$

$$\leq \frac{\|x^{(0)} - x^*\|^2}{2tK}$$

When selecting a proper learning rate, we usually start with a large value like 0.1, then try exponentially lower values like 0.01, 0.001, ... Once we get a good learning rate, where the loss does not increase, we drop the learning rate by 1-2 orders of magnitude and use that instead, since it allow fine-grained learning.