1) Q). we are given that

$$f(t) - p(t) = \frac{1}{n!} f^{(n)}(\theta)\,(t-t_1)(t-t_2)\cdots\cdots(t-t_n)$$

$\theta \in [t_1, t_n]$

also $f^{(n)}(\theta) \le M \quad \forall\, \theta$

Now let $t_i < t_{i+1}$

then $|(t-t_i)(t-t_{i+1})| \le \frac{h^2}{4} \rightarrow (1)$

also, since $t_{i+1} - t_i = h$

$|(t-t_1)(t-t_2)\cdots(t-t_{i-1})| \le |2\cdot 3 \cdots i|$

(Max at $t = t_{i+1}$ for $t \in [t_i, t_{i+1}]$) $\rightarrow (2)$

parallely

$|(t-t_{i+2})(t-t_{i+3})\cdots(t-t_{i+n})| \le |2\cdot 3 \cdots n-i|$

$\le n!$ (Max at $t = t_i$ for $t \in [t_i, t_{i+1}]$) $\rightarrow (3)$

Hence

$$|f(t) - p(t)| = \left|\frac{1}{n!} f^{n}(\theta)(t-t_1)\cdots\cdots t-t_n\right|$$

$$\le \frac{1}{n!} \frac{Mh^2}{4} |(t-t_1)\cdots(t-t_{i-1})|\,|(t-t_{i+2}\cdots t-t_n)| \quad \text{using } (1)$$

$$\le \frac{Mh^2}{4n!}\, i!\,(n-i)! \quad \text{using } (2) \,\&\, (3)$$

$$\le \frac{Mh^2}{4\,{}^{n}C_i}$$

hence proved

COL726

Assignment - 3

K. Laxman
2018CS50408.

Q3) To prove the second order condition for convexity we need to prove

i). $f : \mathbb{R}^n \to \mathbb{R}$ is convex then its Hessian $\nabla^2 f(x)$ is positive semidefinite everywhere

proof:- if f is convex then its Hessian $\nabla^2 f(x)$ is positive semidefinite.

To prove this we will use taylor's theorem with lagrange reminder for a one dimensional case $(n=1)$

$f(x+t) = f(x) + f'(x)t + \frac{1}{2} f''(c) t^2$ where c is number b/w x and $x+t$

Since f is convex we know that

$f(x+t) \geq f(x) + f'(x)t \quad \forall \ x, t \in \mathbb{R}$

consider $t > 0$ and choose small (t) such that $x+t$ and c are both with in domain of f

$\therefore f(x) + f'(x)t + \frac{1}{2} f''(c) t^2 \geq f(x) + f'(x)t$

subtract $f(x)$ and $f'(x)t$ from both sides

$\frac{1}{2} f''(c) t^2 \geq 0$

Since $t > 0$, $\Rightarrow f''(c) \geq 0$ this holds $\forall c$ b/w x and $x+t$

$\lim_{t \to 0} f''(x) \geq 0 \ \forall x$ in domain of f.

So, the hessian $\nabla^2 f(x)$ is positive semidefinite everywhere

Now, we will prove second part.

we assume that the hessian of f is positive semidefinite everywhere

Let x and y be two points in the domain of f and let λ be a scalar b/w 0 and 1.

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-(1/2)\lambda(1-\lambda)(x-y)^T\nabla^2 f(\theta)(x-y)$$

θ is point of line segment b/w x and y

$\because$ since the hessian of f is positive semidefinite everywhere

we have that $\nabla^2 f(\theta) \geq 0$

$\therefore$ therefore $(1/2)\lambda(1-\lambda)(x-y)^T\nabla^2 f(\theta)(x-y) \geq 0$

implies that

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)$$

Thus we have shown that function is convex.

4) a) To prove $x^{(0)}, x^{(1)}, x^{(2)} \ldots$ converges to x^* iff $f'(x^k)$ converge to zero we can use the first order optimality condition for convex functions

The first order optimality condition states that a point x^* is a minimizer of a convex fn f iff $f(x^*)=0$

prove both directions

i) if $x^{(0)}, x^{(1)}, x^{(2)} \ldots$ converges to x^*, then $f'(x^k)$ converges to zero

Given that x^* is unique minimizer of f and f is twice differentiable convex fn

$f'(x^*)=0$

since $x^0, x^1, x^{(2)} \ldots$ converges to x^* (we can use continuity of f & f') to establish $f'(x^k)$ converges to $f'(x^*)$ which is equal to zero

Thus $f'(x^k)$ converges to zero

i) if $f'(x^k)$ converges to zero then $x^0, x^1, x^2 \ldots$ converge to x^*

Lets assume $f'(x^k)$ converge to zero

since f is convex & $f'(x^k) \to 0$ we can use continuity of f'

that $f(x^*)=0$ here x^* is unique minimizer of f

Now by contradiction assume that $x^0, x^1, x^2 \ldots$ does not converge to x^*

This means there exists a subsequence of x^{k_n} that converges to some $\bar{x} \neq x^*$

since $\bar{x} \neq x^*$ and f is convex we have

$f(\bar{x}) > f(x^*) \longrightarrow$ ①

Let's use taylor theorem with lagrange remainder

$f(x) = f(x^*) + f'(c)(x - x^*)$ ($\because c$ is number b/w x and x^*
$f'(x^k) \xrightarrow{\text{converges}}$ to zero)

we can chose sufficiently large k_n such that $f'(c)=0$ $\forall x^{k_n}$

using taylor theorem above

$f(x^{k_n}) = f(x^*) + f'(c)(x^{k_n} - x^*)$

Now $\lim_{n \to \infty} f(x^{k_n}) \to f(\bar{x})$ using continuity of f.

hence we get $f(\bar{x}) = f(x^*)$ so, it is a contradiction ①

Thus the assumption $x^0, x^1, x^2 \ldots$ doesnot converge to x^* must be false & conclude that

$x^0, x^1, x^2 \ldots$ converges to x^*

(b) No, it is not sufficient to guarantee that $x^k \to x^*$ as $k \to \infty$, Let $f(x) = |x|$ and let $x^k = (-1)^k/k$ then we have that $f(x^0) > f(x^1) > f(x^2) \ldots$ but the sequence $\{x^k\}$ doesnot converge to any point.

5 a) we are given that

$$f(x) = \sum_{i=1}^{k} |a_i^T x + b_i|$$

$$f(\lambda x + (1-\lambda) y) = \sum_{i=1}^{k} |\lambda a_i^T x + (1-\lambda) a_i^T y + b_i|$$

$$= \sum_{i=1}^{k} |\lambda a_i^T x + (1-\lambda) a_i^T y + \lambda b_i + (1-\lambda) b_i|$$

$$\leq \sum_{i=1}^{k} |\lambda a_i^T x + \lambda b_i| + |(1-\lambda) a_i^T y) + (1-\lambda) b_i)$$

$$\leq \lambda f(x) + (1-\lambda) f(y)$$ hence proved

Since $f(x) = \sum_{i=1}^{k} |a_i^T x + b_i|$ is not differentiable at various points i.e. where $a_i^T x + b_i = 0$ and gradient descent and newton method both require the function to be differentiable.

5 b) we are given that

$$\hat{f}(x) = \sum_{i=1}^{k} \sqrt{(a_i^T x + b_i)^2 + \epsilon}$$

$$\hat{f}(\lambda x + (1-\lambda) y) = \sum_{i=1}^{k} \sqrt{[\lambda (a_i^T x + b_i) + (1-\lambda)(a_i^T y + b_i)]^2 + \epsilon}$$

$$\leq \sum_{i=1}^{k} \sqrt{[\lambda (a_i^T x + b_i)]^2 + \epsilon(\lambda^2)} \cdot + \sqrt{[(1-\lambda)(a_i^T y + b_i)]^2 + \epsilon(1-\lambda)^2}$$

(using general inequalities)

$$= \lambda \hat{f}(x) + 1-\lambda \hat{f}(y)$$

hence using the general inequalities convex proved

2 Q) The interpolatory quadrature rule in general form is given by

$$\int_{-1}^{1} f(x)dx = \omega_1 f(x_1) + \omega_2 f(x_2)$$

Now we take $f(x)=1$, we get

$$\int_{-1}^{1} 1\,dx = \omega_1 \times 1 + \omega_2 \times 1$$

$$2 = \omega_1 + \omega_2$$

Now we take $f(x) = x$

$$\int_{-1}^{1} x\,dx = \omega_1 x_1 + \omega_2 x_2$$

$$0 = \omega_1 x_1 + \omega_2 x_2$$

In order to have stable algorithm ω_1 & ω_2 must be (+ve)

On solving we get

$$\omega_1 = \frac{2}{1-\frac{x_1}{x_2}}, \quad \omega_2 = \frac{2}{1-\frac{x_2}{x_1}}$$

for $\omega_1 > 0 \Rightarrow \left(1-\frac{x_1}{x_2}\right) > 0 \quad \therefore \frac{x_1}{x_2} < 1$

For $\omega_2 > 0 \Rightarrow \left(1-\frac{x_2}{x_1}\right) > 0 \quad \frac{x_2}{x_1} < 1$

This is only possible when x_1 & x_2 have opposite signs

so, $x_1 \in [-1,0)$ & $x_2 \in (0,1]$

$x_1 \in [0,1]$ or $x_2 \in [-1,0]$

2 b) A point interpolatory quadrature rule has degree 1 and in order to obtain rule with degree 3, we need to choose nodes such that the interpolating polynomial has degree 3.

For $f(x) = 1$

$$\omega_1 + \omega_2 = \int_{-1}^{1} f(x) = 2$$

$f(x) = x$

$$\omega_1 x_1 + \omega_2 x_2 = \int_{-1}^{1} x\,dx = 0$$

$f(x) = x^2$

$$\omega_1 x_1^2 + \omega_2 x_2^2 = \int_{-1}^{1} x^2\,dx = 2/3$$

$f(x) = x^3$

$$\omega_1 x_1^3 + \omega_2 x_2^3 = \int_{-1}^{1} x^3\,dx = 0$$

we can find a solution by taking $\omega_1 = \omega_2 = 1$ and $x_1 = -x_2$

hence

$$x_1^2 + x_2^2 = 2/3 \Rightarrow x_1^2 = 1/3 \Rightarrow x_1 = \pm \frac{1}{\sqrt{3}}$$

hence for nodes $x_1 = \frac{1}{\sqrt{3}}$, $x_2 = \frac{-1}{\sqrt{3}}$

or

$x_1 = -\frac{1}{\sqrt{3}}$ or $x_2 = \frac{1}{\sqrt{3}}$

we get the required solution