HOMEWORK-6

NUMERICAL ALGORITHMS

(1)

To prove $f : R^n \to R$ is convex iff $\tilde{f}(t) = f(x + tv)$ is convex $\forall x, v \in R^n$

\# lets assume f is convex, then to prove that $\tilde{f}(t)$ is convex

$\Rightarrow \tilde{f}(\alpha t_1 + (1-\alpha) t_2) = f(x + (\alpha t_1 + (1-\alpha) t_2) v)$ for any $t_1, t_2 \in domain(\tilde{f})$

$= f(x + \alpha t_1 v + (1-\alpha) t_2 v)$

$= f(\alpha x + \alpha t_1 v + (1-\alpha) x + (1-\alpha) t_2 v)$

$= f(\alpha (x + t_1 v) + (1-\alpha)(x + t_2 v))$

$\leq \alpha f(x + t_1 v) + (1-\alpha) f(x + t_2 v)$

[$\because f$ is assumed to be convex]

$\Rightarrow \tilde{f}(\alpha t_1 + (1-\alpha) t_2) \leq \alpha \tilde{f}(t_1) + (1-\alpha) \tilde{f}(t_2)$

$\Rightarrow \tilde{f}(t)$ is convex if $f(x)$ is convex.

Now, lets assume $\tilde{f}(t)$ is convex

And $f(\alpha x_1 + (1-\alpha)x_2)$ can be written exactly in terms of $\tilde{f}(t)$ as:

$$\tilde{f}(t) = f(x + tv)$$

$$\tilde{f}(t) = f(x_1 + t(x_2 - x_1)) \quad \text{if } x = x_1 \text{ and } v = x_2 - x_1$$

$$\Rightarrow \tilde{f}(0) = f(x_1)$$

$$\tilde{f}(1) = f(x_2)$$

and
$$\tilde{f}(1-\alpha) = f(x_1 + (1-\alpha)(x_2 - x_1))$$

$$= f(x_1 + (1-\alpha)x_2 - (1-\alpha)x_1)$$

$$\Rightarrow \tilde{f}(1-\alpha) = f(\alpha x_1 + (1-\alpha)x_2)$$

$$\Rightarrow f(\alpha x_1 + (1-\alpha)x_2) = \tilde{f}(1-\alpha)$$

$$= \tilde{f}(\alpha \cdot 0 + (1-\alpha)\cdot 1)$$

$$\leq \alpha \tilde{f}(0) + (1-\alpha)\tilde{f}(1)$$

$$\boxed{f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)},$$

$\Rightarrow$ f is convex if $\tilde{f}$ is convex

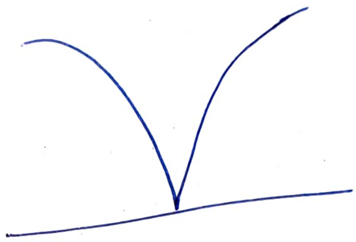
$\Rightarrow$ $f: R^n \to R$ is convex if and only if $\tilde{f}(t) = f(x + tv)$ is convex.

⇒ If we restrict the lines only to the co-ordinate axes, Then this condition does not holds ie.,

$f(x)$ can be non convex even if $\tilde{f}(t) = f(x+te_i)$ is convex.

⇒ An example of such a function can be the quasiconvex function given in the slides

ie, $f(x) = \sqrt{|x|}$



⇒ In this function $\tilde{f}(t)=f(x+te_i)$ is convex but function itself is not convex.

⇒ $\tilde{f}(t)= f(x+te_i)$ is convex because from any point, ~~if w the~~ line from it parallel to coordinates axis ~~∈ Domain~~ belong to the set but the two points which are not considered joined by lines parallel to coordinate axis are not like

(x, y) joined by line which proves that $\sqrt{|x|}$ itself is not convex, since

$f(\theta x+(1-\theta)y) > \theta f(x)+(1-\theta)f(y)$

②

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

let $f_i(x) = |a_i^T x + b_i|$

Then $f(x) = \sum_{i=1}^{k} f_i(x)$

$\Rightarrow$ Prooving each $f_i(x)$ is convex:

$\Rightarrow$ ~~$f_i(x)$~~ Every (input) domain $(f_i(x)) = R^n$

$\Rightarrow$ Input set is convex

and $f_i(\alpha x_1 + (1-\alpha)x_2) = |a_i^T(\alpha x_1 + (1-\alpha)x_2) + b_i|$

$= |\alpha a_i^T x_1 + (1-\alpha) a_i^T x_2 + \alpha b_i + (1-\alpha) b_i|$

$= |\alpha(a_i^T x_1 + b_i) + (1-\alpha)(a_i^T x_2 + b_i)|$

$\leq |\alpha(a_i^T x_1 + b_i)| + |(1-\alpha)(a_i^T x_2 + b_i)|$

$[\because |x_1 + x_2| \leq |x_1| + |x_2|]$

$\leq \alpha|a_i^T x_1 + b_i| + (1-\alpha)|a_i^T x_2 + b_i|$

$(\because \alpha \geq 0 \ \& \ \alpha \leq 1)$

$\Rightarrow f_i(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f_i(x_1) + (1-\alpha) f_i(x_2)$

$\Rightarrow$ Each $f_i(x)$ is convex.

$\Rightarrow f(x) = f_1(x) + f_2(x) + \dots + f_k(x)$

Again $\mathrm{Dom}(f) \in R^n \Rightarrow$ Input set is convex and $f(\alpha x_1 + (1-\alpha) x_2)$

$$= f_1(\alpha x_1 + (1-\alpha) x_2) + \cdots + f_K(\alpha x_1 + (1-\alpha) x_2)$$

$$\leq \alpha f_1(x_1) + (1-\alpha) f_1(x_2) + \cdots + \alpha f_K(x_1) + (1-\alpha) f_K(x_2)$$

$$\leq \alpha [f_1(x_1) + f_2(x_1) + \cdots + f_K(x_1)] + (1-\alpha)[f_1(x_2) + f_2(x_2) + \cdots + f_K(x_2)]$$

$$\leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$\Rightarrow$ $f(x)$ is also convex

[$\because$ Sum of convex functions is convex as proved above].

However $f(x)$ cannot be effectively minimized by Gradient Descent and Newton's method.

$\rightarrow$ The function $f(x) = \sum_{i=1}^{K} |a_i^T x + b_i|$ is not differentiable when, x = solution of $a_i^T x + b = 0$ (for each $i = 1 \ldots K$)

i.e., if $a_i^T x + b = 0$, then $f_i(x)$ is not differentiable and hence $f(x)$ is not differentiable.

$\Rightarrow$ When minimizing using gradient descent we might not be able to take the exact step $x^t = x + t\Delta x$ since Δx may not be possible to calculate for some

x and thus we might be approximating Δx here by some value which is not as effective.

Thus, gradient descent won't effectively minimize $f(x)$.

→ Similarly, the second derivative for each $f_i(x)$ doesn't exist or vanishes and thus the second derivative for $f(x)$ vanishes.

Thus, Newton's Method is also not able to effectively minimize this $f(x)$.

[$\because \Delta x = -H(x)^{-1} g(x)$] and $H(x)$ vanishes here]

For the smoothed function

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

consider each $\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + \epsilon}$

Then $\hat{f}(x) = \sum_{i=1}^{K} \hat{f}_i(x)$

⇒ For $\hat{f}_i(x)$, $x \in \mathbb{R}^n$ ⇒ Domain is convex.

& $\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + \epsilon}$

(can be written in terms of 2-norm)

$$\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + (\sqrt{\epsilon})^2} = \left\| \begin{matrix} a_i^T x + b_i \\ \sqrt{\epsilon} \end{matrix} \right\|_2$$

Then $\hat{f}_i(\alpha x_1 + (1-\alpha)x_2) = \left\| \dfrac{a_i^T(\alpha x_1 + (1-\alpha)x_2) + b_i}{\sqrt{E}} \right\|_2$

$$= \left\| \frac{\alpha a_i^T x_1 + (1-\alpha) a_i^T x_2 + \alpha b_i + (1-\alpha) b_i}{\alpha\sqrt{E} + (1-\alpha)\sqrt{E}} \right\|_2$$

$$= \left\| \frac{\alpha(a_i^T x_1 + b_i) + (1-\alpha)(a_i^T x_2 + b_i)}{\alpha\sqrt{E} + (1-\alpha)\sqrt{E}} \right\|_2$$

$$\leq \left\| \frac{\alpha(a_i^T x + b_i)}{\alpha\sqrt{E}} \right\|_2 + \left\| \frac{(1-\alpha)(a_i^T x_2 + b_i)}{(1-\alpha)\sqrt{E}} \right\|_2$$

(following $\|X+Y\| \leq \|X\| + \|Y\|$ and $\|cX\| = |c|\,\|X\|$)

$$\cancel{\leq |\alpha| \left\| \frac{a_i x_1 + b_i}{\sqrt{E}} \right\| + |(1-\alpha)| \left\| (1-\alpha) \right\|}$$

$$\leq |\alpha| \left\| \frac{(a_i^T x_1 + b_i)}{\sqrt{E}} \right\|_2 + |(1-\alpha)| \left\| \frac{(a_i^T x_2 + b_i)}{\sqrt{E}} \right\|_2$$

$$\boxed{\Rightarrow \hat{f}_i(\alpha x_1 + (1-\alpha)x_2) \leq \alpha \hat{f}_i(x_1) + (1-\alpha)\hat{f}_i(x_2)}$$

($\because$ Both α, $(1-\alpha) \geq 0$)

$\Rightarrow \hat{f}_i(x)$ is convex

$\Rightarrow \hat{f}(x) = \sum_{i=1}^{k} \hat{f}_i(x)$ is also convex since sum of convex functions is convex as proved previously.

③ (a)

$$f(x) = \sum_{(i,j) \in E} \|x_i - x_j\|_2^p$$

→ The gradient for each term in the summation can be calculated as :

$$\frac{\partial \|x_i - x_j\|_2^p}{\partial x_i} = p\,\|x_i - x_j\|_2^{p-1} \cdot \frac{\cancel{2}(x_i - x_j)}{\cancel{2}\sqrt{(x_i - x_j)^T(x_i - x_j)}}$$

$$= p\,\|x_i - x_j\|_2^{p-2} \cdot (x_i - x_j) \quad \text{[Using chain rule)}$$

$$\frac{\partial(\|x_i - x_j\|_2^p)}{\partial x_j} = -p\,\|x_i - x_j\|_2^{p-2}(x_i - x_j)$$

⇒ The gradient of f can be calculated in the same way by adding the gradient of each term x_i, x_j to the gradient vector.

[Note that gradient $(x_i) = 0$ if $i <= m$]
(since they are the fixed nodes).

⇒ Implemented Gradient Descent with line search and the graphs are as follows :