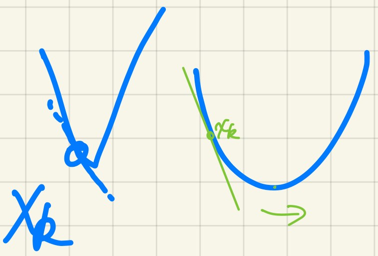
Question 1:

why not converge?



increase μ parameter

if $f(x_{k-1}) - f(x_k) \approx 0$

Add Another termination condition

Question 8:

pesudo-huber function:

$$\varphi_\mu(x) = \sum_{i=1}^{n} \left((\mu^2 + |x_i|^2)^{1/2} - \mu \right)$$

Convex conjugate:

$$f^*(y) := \max_{x \in \mathbb{R}^n} y^T x - f(x)$$

$$= \max_x x^T y - \|x\|_1$$

$$= \max_x \sum_{i=1}^n x_i y_i - \sum_{i=1}^n |x_i|$$

$$= \max \sum_{i=1}^n x_i y_i - |x_i|$$

$$= \sum_{i=1}^n \underline{\max\ x_i y_i - |x_i|}$$

$$\max\ x_i y_i - |x_i| = \max_{x_i} x_i y_i - \mathrm{Sign}(x_i) \cdot x_i$$

$$= \max_{x_i} x_i \left(y_i - \mathrm{Sign}(x_i) \right)$$

y_i	x_i	max
$y_i > 1$	$x_i = +\infty$	$+\infty$
$y_i < -1$	$x_i = -\infty$	$+\infty$
$-1 \le y_i \le 1$	$x_i = 0$.	Zero.

$$\max_{x_i} x_i y_i - |x_i| = \begin{cases} 0 & |y_i| \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

In order to maximize $f_\mu(x)$. we need to minimize $f^*(y)$, take $|y_i| < 1$ and $f^*(y_i) = 0$

$$f_\mu(x) = \max_{y \in \mathbb{R}^n} x^T y - f^*(y) - \mu d(y)$$

$$d(y) = \frac{1}{2}\|y\|_2^2 = \frac{1}{2} \sum_{i=1}^n y_i^2$$

$$f_\mu(x) = \max_{y \in \mathbb{R}^n} x^T y - \max_{x \in \mathbb{R}^n} y^T x - f(x) - \frac{\mu}{2} \sum_{i=1}^n y_i$$

$$f_\mu(x) = \max\ x^T y - 0 - \mu d y$$

$$= \max \sum_{i=1}^n x_i y_i - \frac{\mu}{2} \sum_{i=1}^n y_i^2 \qquad |y_i| < 1.$$

$$= \sum_{i=1}^n \max_{|y_i| < 1} x_i y_i - \frac{\mu}{2} y_i^2$$

Gradients w.s.t. y_i

$$\nabla f_\mu(x) = x_i - \mu y_i = 0$$

$$y_i = \frac{x_i}{\mu} \Rightarrow \begin{cases} y_i = 1 & \text{if } |x_i| > \mu \\ y_i = \frac{x_i}{\mu} & \text{if } |x_i| \le \mu \end{cases}$$

but we have a constrain $|y_i| \le 1$

Therefore.

$$f_\mu(x) = \max\ y_i x_i - \frac{\mu}{2} y_i^2 = \begin{cases} \frac{\mu}{2} y_i^2 & \text{if } |x_i| \le \mu \\ |x_i| - \frac{1}{2} & \text{if } |x_i| > \mu \end{cases}$$

Question 9.

$$d(y) = \sum_{i=1}^{n} 1-\sqrt{1-y_i^2} \qquad |y_i| \le 1 \quad \forall i.$$

$$f_\mu(x) = \max_{y\in\mathbb{R}^n} \; x^T y - f^*(y) - \mu d(y)$$

$$f^*(y) = \max_{x\in\mathbb{R}^n} y^T x - f(x) = \sum_{i=1}^{n} \max\left(x_i y_i - x_i\right), \quad \max(x_i y_i - |x_i|) = \begin{cases} 0 & |y_i| \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$f_\mu(x) = \max\, x y^T - f^*(y) - \mu d(y)$$

$$= \max\left(x_i y_i\right) - \mu\left(\sum_{i=1}^{n} 1 - \sqrt{1-y_i^2}\right) \qquad |y_i| \le 1$$

$$= \underline{\max(x_i y_i)} - \underline{\mu \sum_{i=1}^{n} 1} + \underline{\mu \sum_{i=1}^{n} (1-y_i^2)^{1/2}}.$$

The gradient of f_μ wrt. y_i

$$\frac{y^2}{1-y^2} = \frac{x^2}{\mu^2} \qquad y = \frac{x^2}{\mu^2+x^2}$$

$$\nabla f_\mu(x) = x_i - \mu y_i (1-y_i^2)^{-1/2} = 0$$

$$y_1 = \frac{x}{(\mu^2+x^2)^{1/2}} \quad y_2 = -\frac{x}{(\mu^2+x^2)^{1/2}}. \Leftarrow |y|<1 \text{ guaranteed.}$$

$$f(x_i) = x_i \frac{x_i}{(\mu^2+x_i^2)^{1/2}} - \mu + \mu\left(1 - \frac{\mu}{\sqrt{\mu^2+x_i^2}}\right)$$

$$= \frac{x_i^2}{(\mu^2+x_i)^{1/2}} + \frac{\mu^2}{(\mu^2+x_i)^{1/2}} - \mu$$

$$= \frac{x_i^2+\mu^2}{(\mu^2+x_i^2)^{1/2}} - \mu \Leftarrow \text{inside } \Sigma.$$

Sum for i from 1 to n.

$$\varphi_\mu(x) = \sum_{i=1}^{n}\left(\left(|x_i|^2+\mu^2\right)^{1/2} - \mu\right)$$

Question 10.

Show $\varphi_\mu(x) = \sum_{i=1}^{n}\left((\mu^2+|x_i|^2)^{1/2}-\mu\right)$ is convex but not strongly convex.

Assume $\forall x_1, x_2 \in X$ and $t\in[0,1]$

Convex

$$\varphi_u(tx_1+(1-t)x_2) = \sum_{i=1}^{n}\left((\mu^2+(tx_{1i}+(1-t)x_{2i})^2)^{1/2}-\mu\right)$$

$$=\sum_{i=1}^{n}\left(\left(\mu^2+t^2x_{1i}^2+(1-t)^2x_{2i}^2+2t(1-t)x_{1i}x_{2i}\right)^{1/2}-\mu\right) \quad \text{— } 1$$

$$t\varphi_u(x_1)+(1-t)\varphi_u(x_2)$$

$$=t\cdot\sum_{i=1}^{n}\left((\mu^2+x_{1i}^2)^{1/2}-\mu\right)+(1-t)\sum_{i=1}^{n}\left((\mu^2+x_{2i}^2)^{1/2}-\mu\right)$$

$$=\sum_{i=1}^{n}\left(t(\mu^2+x_{1i}^2)^{1/2}+(1-t)(\mu^2+x_{2i}^2)^{1/2}-\mu\right) \quad \text{— } 2$$

Since $(ax+by)^2 \geq (ax)^2+by$

where $a=t$. $x=(\mu^2+x_{1i}^2)^{1/2}$

$b=1-t$. $y=(\mu^2+x_{2i}^2)^{1/2}$

Note. $a\geq 0$. $b\geq 0$. $x\geq 0$. $y\geq 0$. guarenteed for $t\in(0,1)$

$1^2=(ax)^2+(by)^2$

$2^2=(ax+by)^2$

$\Rightarrow 1^2 < 2^2$ Such that $1 < 2$

$\Rightarrow \varphi_u(tx_1+(1-t)x_2) < t\varphi_u(x_1)+(1-t)\varphi_u(x_2)$. $\Rightarrow$ $\varphi_u(x)$ is convex

---

$$\nabla_{x_i}\varphi_u(x) = \frac{x_i}{(\mu^2+x_i^2)^{1/2}}.$$

Assume $y>x$.

$$f(y)-f(x)-\nabla f(x)^T(y-x)$$

$$\sum_{i=1}^{n}\left((\mu^2+|y_i|^2)^{1/2}-\mu\right)-\sum_{i=1}^{n}\left((\mu^2-|x_i|^2)^{1/2}-\mu\right)-\sum_{i=1}^{n}\frac{x_i}{(\mu^2+x_i^2)^{1/2}}(y_i-x_i)$$

$$=\sum_{i=1}^{n}\left[(\mu^2+|y_i|^2)^{1/2}-(\mu^2+|x_i|^2)^{1/2}-x_i(\mu^2+x_i^2)^{-1/2}(y_i-x_i)\right].$$

multiply by $(y_i^2+\mu^2)^{1/2}+(x_i^2+\mu^2)^{1/2}$

$$\nabla\varphi_u(x)(y-x)\cdot\left((y^2+u^2)+(x^2+\mu^2)\right)^{1/2}=\sum_{i=1}^{n}x_i(y_i-x_i)+x_i(y_i-x_i)(y_i^2+\mu^2)^{1/2}(x_i^2+\mu^2)^{-1/2}$$

$$\varphi_u(y)-\varphi_u(x)\left[(y^2+u^2)^{1/2}+(x^2+u^2)^{1/2}\right]=\sum_{i=1}^{n}(y_i^2+\mu^2)-(x_i^2-\mu^2)=\sum_{i=1}^{n}(y_i^2-x_i^2)$$

$$\varphi_u(y)-\varphi_u(x)-\nabla_u f(x)(y-x)=\sum_{i=1}^{n}y_i^2-x_i^2-x_iy_i+x_i^2-x_i(y_i-x_i)(y_i^2+\mu^2)^{1/2}(x_i^2+\mu^2)^{-1/2} \quad \text{— } 3$$

$$=\sum_{i=1}^{n}y_i(y_i-x_i)-\frac{(y_i^2+\mu^2)^{1/2}}{(x_i^2+\mu^2)^{1/2}}x_i(y_i-x_i).$$

$$=\sum_{i=1}^{n}(y_i-x_i)\boxed{\left(y_i-\frac{(y_i^2+\mu^2)^{1/2}}{(x_i^2+\mu^2)^{1/2}}x_i\right)}$$

Assume $y_i > x_i$

$$\Rightarrow y_i^2 - \frac{y_i^2 x_i^2 + \mu^2 x_i^2}{x_i^2+\mu^2} = \frac{y_i^2\mu^2 - x_i^2\mu^2}{x_i^2+\mu^2} = \frac{\mu^2}{x_i^2+\mu^2}\underline{(y_i^2 - x_i^2)} < 0.$$

if $0 > y_i > x_i$

$y_i^2 - x_i^2 < 0$. $y_i - x_i > 0$.

if $y_i^2 < x_i^2$, then ① < ②

then: $\varphi_u(y) - \varphi_u(x) - \nabla f(x)(y-x) < 0$.

Not $\varphi_u(y) \geq \varphi_u(x) + \nabla f(x)(y-x) + \frac{\mu}{2}\|y-x\|_2^2$

$\Rightarrow$ $\varphi_u(x)$ is not strongly convex.

Question 11.

f is strongly convex

$$f^*(y) = \max_{y \in \text{dom} f^*} x^T y - f^*(y) \Rightarrow \nabla f(x) = \underset{y \in \text{dom} f}{\text{argmax}} \; x^T y - f^*(y)$$

we have $f^*(y) = \sup_{u \in \text{dom} f} y^T u - f(u) = \max_{u \in \text{dom} f} y^T u - f(u)$

because of the convexity of f

lets $u^* = x$ is the global minimizer

$$0 \in \partial[y^T x - f(x)] = y - \partial f(x) \iff y \in \partial f(x)$$

$$\Rightarrow f^*(y) = y^T x - f(x)$$

for any v. we have:

$$f^*(v) = \sup_u v^T u - f(u) \geq v^T x - f(x) \geq v^T x - y^T x + y^T x - f(x)$$

$$= f^*(y) + (y - v)^T x$$

which

$$f^*(v) \geq f^*(y) + (v - y)^T x$$

By definition of subgradients. we have $x \in \partial f^*(y)$

By second conjugate function's property

$$f^{**}(x) = f(x)$$

We proved:

$$y \in \partial f(x) \Rightarrow f(x) + f^*(x) = x^T y \Rightarrow x \in \partial f^*(y) \Rightarrow y \in \partial f(x).$$

So that $\underline{x \in \partial f^*(y) \Rightarrow \partial f^{**}(x) = \partial f(x).}$ #

$\max_{y \in \text{dom} f^*} x^T y - f^*(y)$ has an unique maximizer x.

from #. $x \in \partial f^*(y)$. since its is the <u>unique</u>. maximizer.

$\partial f^*(y)$ has only one elements.

$$\partial f^*(y) = \{x\} = \{\nabla f^*(y)\}.$$

$\Rightarrow$ f^* is differentiable and

$$\nabla f(x) = \underset{y \in \text{dom} f^*}{\text{argmax}} \; x^T y - f^*(y)$$

Question 12:

$f(x)$ is δ-strongly-convex: we have (by definition)

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \delta \|x-y\|_2^2$$

①

$$(\nabla f(x) - \nabla f(y))^T (\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))) \geq \mu \|\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))\|_2^2$$

②

$$\|\nabla f(x) - \nabla f(y)\|_2 \cdot \|\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))\|_2 \geq (\nabla f(x) - \nabla f(y))^T (\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y)))$$

by ① and ②

$$\|\nabla f(x) - \nabla f(y)\|_2 \cdot \|\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))\| \geq \mu \|\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))\|_2^2$$

$$\frac{1}{\mu} \|\nabla f(x) - \nabla f(y)\|_2 \geq \|\nabla f^*(\nabla f(x)) - \nabla f^*(\nabla f(y))\|_2$$

$\Rightarrow$ $\nabla f(x)$ is continuous with Lipschitz Constant $1/\delta$

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \delta \|x-y\|_2^2 \quad \text{(multiplied both side)}$$

And. $f^{**}(x) = (f^*(y))^* = f(x)$.

$$f^{**}(x) = \max_{y \in \mathbb{R}^n} x^T y - f^*(y) \geq x^T y.$$

Question 13

dom $f^* \subseteq R^n$ is closed and bounded

$D = \max_{y=c} d(y)$. bounded $D < +\infty$ dom $f^* \subseteq$ dom d.

$f(x) = y^T x + a$.

$\bar{D} = \max_{y \in \text{dom } d} d(y) < +\infty$ because f^* is bounded and dom $f^* \subset$ dom d.

$$f^* = \max_{x \in \text{dom } d} y^T x - f(y) = \sup_{y \in \text{dom } d} (y^T x - f(y))$$

$$f(x) = \max_{x \in \text{dom } d} y^T x - f^*(y) = \sup_{y \in \text{dom } d} (y^T x - f^*(y))$$

$$f_\mu(x) = \max_{y \in \text{dom } d} y^T x - f^*(y) - \mu d(y).$$

$$f_\mu(x) = \sup_{y \in \text{dom } d} y^T x - f^*(y) - \mu d(y)$$

$$\geq \sup_{y \in \text{dom } d} y^T x - f^*(y) - \mu D.$$

$$= f(x) - \mu D.$$

proof of lower bound

$$f_\mu(x) \leq \sup_{y \in \text{dom } d} y^T x - f^*(y) - \mu d(y) + \underline{\mu d(y)}$$

$$= \sup_{y \in \text{dom } d} y^T x - f^*(y) = f(x).$$

proof of upperbound.

Question 14:

$\phi_0(x) = f(x_0) + \frac{L}{2}\|x - x_0\|_2^2$. $\lambda_k \le \frac{c}{k^2}$, $c > 0$. by definition of $\phi_k(x)$, it is obvious that $f(x_k) - f^* \le \phi_k(x_k) - f^*$

$$
\begin{aligned}
f(x_k) - f^* &\le \phi_k(x_k) - f^* \\
&\le (1-\lambda_k) f(x^*) + \lambda_k \left(f(x_0) + \frac{L}{2}\|x^* - x_0\|_2^2 \right) - f^* \\
&= f(x^*) - \lambda_k f(x^*) + \lambda_k f(x_0) + \frac{\lambda_k L}{2}\|x^* - x_0\|_2^2 - f^* \\
&= \lambda_k \left[-f(x^*) - f(x_0) + \frac{L}{2}\|x^* - x_0\|_2^2 \right] \\
&\le \frac{c}{k^2} \left(-f^* + f(x_0) + \frac{L}{2}\|x^* - x_0\|_2^2 \right) \\
&\le \frac{1}{k^2} \cdot c \left(-f^* - f(x_0) + \frac{L}{2}\|x^* - x_0\|_2^2 \right)
\end{aligned}
$$

Question 15

· We have proved that gradient descent for δ strongly convex function $f(x_k) - f^* \le (1-\frac{\delta}{L})(f(x)-f^*)$

we have proved that $f(x_k) - f^* = O(\frac{1}{k^2})$

For accelerated gradient descent method:

$$f(x_k) - f^* \le \left(1 - \sqrt{\frac{\delta}{L}}\right)^k (f(x_0) - f^*)$$

$$e^{-x} \ge e^x \longrightarrow \quad \le e^{-k\sqrt{\frac{\delta}{L}}}(f(x_0) - f^*)$$

let $f(x_k) - f^* = \varepsilon$

$$e^{k\sqrt{\frac{\delta}{L}}}[f(x_0) - f^*] \le \varepsilon$$

$$-k\sqrt{\frac{\delta}{L}} + \log[f(x_0) - f^*] \le \log \varepsilon$$

$$k \ge \left(\frac{\delta}{L}\right)^{-1/2} \cdot [\log(f(x_0) - f^*) - \log \varepsilon]$$

$$= \sqrt{\frac{L}{\delta}} \log\left(\frac{f(x_0) - f^*}{\varepsilon}\right)$$

$$\Rightarrow k = O\left(\sqrt{\frac{L}{\delta}} \log \frac{1}{\varepsilon}\right).$$