1) $\text{prox}_f(\bar{x}) = \underset{x \in \mathbb{R}^d}{\text{argmin}} \left\{ f(x) + \frac{1}{2}\|x-\bar{x}\|_2^2 \right\}$

a) $f(x) = \begin{cases} 0 & \text{if } \|x\|_\infty \le 1 \\ +\infty & \text{otherwise} \end{cases}$, so when $\|\bar{x}\|_\infty \le 1$ $\text{prox}_f(\bar{x}) = \bar{x}$

when $\|\bar{x}\|_\infty > 1$, we need to minimize $\frac{1}{2}\|x-\bar{x}\|_2^2$ and let $\|x\|_\infty \le 1$

so if $-1 \le \bar{x}_i \le 1$ $\quad x_i = \bar{x}_i$

if $\bar{x}_i < -1$ $\quad x_i = -1$

if $\bar{x}_i > 1$ $\quad x_i = 1$

b) $f(x) = \alpha \|x\|_3^3 = \sum_{i=1}^{d} \alpha |x_i|^3$

assume $\alpha > 0$ since if $\alpha < 0$ and $|x_i|^3$ grows faster that $|x_i|^2$
there is no minimizer,

then, $\text{prox}_f(\bar{x}) = \underset{x}{\text{argmin}} \left\{ \sum_{i=1}^{d} \alpha |x_i|^3 + \frac{1}{2}\|x-\bar{x}\|_2^2 \right\}$

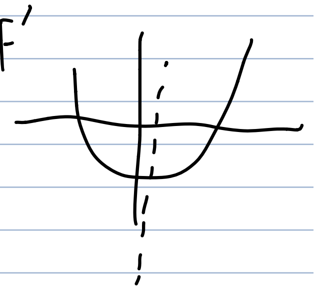
$= \underset{x}{\text{argmin}} \left\{ \sum_{i=1}^{d} \alpha |x_i|^3 + \frac{1}{2}|x_i - \bar{x}_i|^2 \right\}$

when $\bar{x}_i > 0$ then x_i^* need to ≥ 0 and $x_i^* \le \bar{x}_i$

then, $\alpha |x_i|^3 + \frac{1}{2}|x_i - \bar{x}_i|^2 = \alpha x_i^3 + \frac{1}{2}(x_i - \bar{x}_i)^2$ $\quad (\alpha > 0)$

let $F_i(x) = \alpha x^3 + \frac{1}{2}(x-\bar{x}_i)^2$ $\quad \alpha, \bar{x}_i > 0$

F'



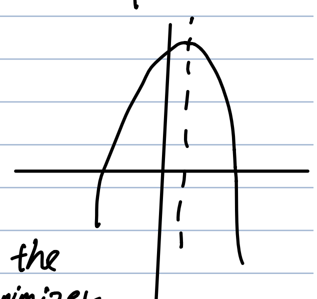
$F_i'(x) = 3\alpha x^2 + x - \bar{x}_i$ $\quad F_i''(x) = 6\alpha x + 1$ $\quad$ when $F_i'(x) = 0$ $\quad x = \frac{-1 \pm \sqrt{1+12\alpha\bar{x}_i}}{6\alpha}$,

since $\alpha > 0$ and $1 + 12\alpha\bar{x}_i > 1$, so $x = \frac{-1+\sqrt{1+12\alpha\bar{x}_i}}{6\alpha}$ is the minimizer.

$x_i^* = \frac{-1+\sqrt{1+12\alpha\bar{x}_i}}{6\alpha} < \bar{x}_i$ $\quad$ since $1 + 12\alpha\bar{x}_i \le (1+6\alpha\bar{x}_i)^2$

$a^2 + 2ab \le (a+b)^2$

F'



if $\bar{x}_i < 0$, then, $\bar{x}_i \le x_i^* \le 0$ then, $F_i(x) = -\alpha x^3 + \frac{1}{2}(x-\bar{x}_i)^2$

$F_i'(x) = -3\alpha x^2 + x - \bar{x}_i$ when $F_i'(x) = 0$, $x = \frac{1 \pm \sqrt{1-12\alpha\bar{x}_i}}{6\alpha}$ $\quad x = \frac{1-\sqrt{1-12\alpha\bar{x}_i}}{6\alpha}$ is the minimizer

$x = \frac{1-\sqrt{1-12\alpha\bar{x}_i}}{6\alpha} > \bar{x}_i$ Since $1 - 6\alpha\bar{x}_i > \sqrt{1-12\alpha\bar{x}_i}$

when $\alpha > 0$ $\bar{x}_i < 0$, $(a-b)^2 \geq a^2 - 2ab$

to conclusion: when $\bar{x}_i > 0$ $\quad x_i = \frac{-1+\sqrt{1+12\alpha\bar{x}_i}}{6\alpha}$

when $\bar{x}_i < 0$ $\quad x_i = \frac{1-\sqrt{1-12\alpha\bar{x}_i}}{6\alpha}$

2) for μ-strongly convex:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|_2^2$$

so $\frac{\mu}{2}\|y-x\|_2^2 \leq f(y) - f(x) - \nabla f(x)^T (y-x)$ let "y" $= x$ "x" $= \bar{x}$ in our case.

since $f(x) < f(\bar{x})$. $f(x) - f(\bar{x}) < 0$

then, $\frac{\mu}{2}\|x-\bar{x}\|_2^2 \leq f(x) - f(\bar{x}) - \nabla f(\bar{x})^T (x-\bar{x}) < \nabla f(\bar{x})^T (\bar{x}-x) \leq \|\nabla f(\bar{x})\|_2 \cdot \|x-\bar{x}\|_2$

so $\frac{\mu}{2}\|x-\bar{x}\|_2 \leq \|\nabla f(\bar{x})\|_2$, $\Rightarrow \|x-\bar{x}\|_2 \leq \frac{2}{\mu}\|\nabla f(\bar{x})\|_2$

since $f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|_2^2$

$$= f(x) + \frac{\mu}{2} \cdot \left(\frac{2}{\mu}\nabla f(x)^T (y-x) + \|y-x\|_2^2 + \frac{1}{\mu}\|\nabla f(x)\|_2^2\right) - \frac{1}{2\mu}\|\nabla f(x)\|_2^2 \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|_2^2$$

replace "y" to x "x" to $\bar{x}$

we have $f(x) \geq f(\bar{x}) - \frac{1}{2\mu}\|\nabla f(\bar{x})\|_2^2$

2b) from (a) we know for $f(x) < f(\bar{x})$, we have $\|x-\bar{x}\|_2 \leq \frac{2\|\nabla f(x)\|}{\mu}$

Which means all the $x \in \mathbb{R}^n$ and $f(x) < f(\bar{x})$ are in a compact area.

Since f is a continuously differentiable function then, f is "smooth"

since $f(x) \geq f(\bar{x}) - \frac{1}{2\mu}\|\nabla f(\bar{x})\|_2^2$, then the right hand side could be a lower bound

since all the x that makes $f(x) < f(\bar{x})$ is in a compact area, and $f(x)$ is smooth and has a lower bound, then $\min f(x)$, $\|x-\bar{x}\|_2 \leq \frac{2\|\nabla f(\bar{x})\|_2}{\mu}$ has at least one solution

Hence, there must exist a minimizer of f

2c) if the minimizer is not unique.

then, there are $f(x_1^*) = f(x_2^*)$ and $x_1^* \neq x_2^*$, $\forall x \in \mathbb{R}^n$: $f(x) \geq f(x_1^*) = f(x_2^*)$

from HW2 Q3(a), we have $f(x) \geq f(x^*) + \frac{\mu}{2}\|x - x^*\|_2^2$.

then, $f(x_1^*) \geq f(x_2^*) + \frac{\mu}{2}\|x_1^* - x_2^*\|_2^2$ since $x_1^* \neq x_2^*$, then $f(x_1^*) > f(x_2^*)$

then x_1^* is not a minimizer.

So the minimizer is unique

3 a) first we need to prove : $\partial f(x) + \partial h(x) \subseteq \partial (f+h)(x)$

Since $f(y) \geq f(x) + g_f^T (y-x)$, $h(y) \geq h(x) + g_h^T (y-x)$

then $f(y) + h(y) \geq f(x) + h(x) + (g_f^T + g_h^T)(y-x)$

so $g_f + g_h \in \partial (f+h)$

let $L = \sum_{i=1}^{n} \max\{0, 1 - y_i \cdot x_i^T w\}$ $\quad h = \frac{\lambda}{2}\|w\|_2^2$

$\partial_w L = \sum_{i=1}^{n} g_i \quad g_i = \begin{cases} 0 & \text{if } y_i \cdot x_i^T \cdot w > 1 \\ -y_i \cdot x_i^T & \text{otherwise.} \end{cases}$

$\partial_w h = \lambda w$

so a subgradient of f can be computed as

$$\sum_{i=1}^{n} g_i + \lambda w \in \partial f(w) \quad \text{where } g_i = \begin{cases} 0 & \text{if } y_i \cdot x_i^T \cdot w > 1 \\ -y_i \cdot x_i^T & \text{otherwise.} \end{cases}$$

b. $$\text{prox}_{\alpha f}(0) = \underset{w}{\arg\min} \left\{ f(w) + \frac{1}{2\alpha} \| w - \vec{0} \|_2^2 \right\}$$

$$= \underset{w}{\arg\min} \left\{ \sum_{i=1}^{n} \max\{0, 1 - y_i \cdot x_i^T w\} + \left(1 + \frac{1}{2\alpha}\right) \|w\|_2^2 \right\}$$

which is another SVM problem

Q4 For any x, verify that a subgradient of f at x is given by

$$g(x) := A^T(Ax - b) + \gamma\ sign(x)$$

a) let $L(x) = \frac{1}{2}\|Ax-b\|_2^2$ $h(x) = \gamma\|x\|_1$, so $f(x) = L(x) + h(x)$

$\partial f = \partial L + \partial h$

$\nabla L(x) = A^T(Ax-b)$, so $g_L(x) = \nabla L(x) = A^T(Ax-b)$

in class, we have shown that subgradient of $|x|$ is $\begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$

in Q3(a) I proved $\partial f(x) + \partial h(x) \subseteq \partial(f+h)(x)$

so subgradient of f is: $A^T(Ax-b) + \gamma \cdot sign(x) = g(x)$

b)

```
iteration:  95
objective value is:  11.145629487256196

iteration:  96
objective value is:  11.158322500001669

iteration:  97
objective value is:  11.180967858346426

iteration:  98
objective value is:  11.143029145221693

iteration:  99
objective value is:  11.175554311266895

iteration:  100
objective value is:  11.126944245014103
```

```
In [12]: num = len(np.where(x_kn1 == 0)[0])
         print("number of zeros: ",num)

number of zeros:  0
```

c)

```
iteration:  95
objective value is:  10.335051255078007

iteration:  96
objective value is:  10.329917492481362

iteration:  97
objective value is:  10.324863174433856

iteration:  98
objective value is:  10.319959295630866

iteration:  99
objective value is:  10.315140200605923

iteration:  100
objective value is:  10.31040479095528
```

```
In [16]: num = len(np.where(x_kn1 == 0)[0])
         print("number of zeros: ",num)

number of zeros:  810
```

d)

```
iteration:  95
objective value is:  9.87462731576255

iteration:  96
objective value is:  9.877119774739894

iteration:  97
objective value is:  9.87433786714285

iteration:  98
objective value is:  9.874020858460415

iteration:  99
objective value is:  9.870592008789627

iteration:  100
objective value is:  9.869430185882859
```

```
In [19]: num = len(np.where(x_kn1 == 0)[0])
         print("number of zeros: ",num)

number of zeros:  896
```