Assignment 2

1. a. $\mathbb{1}(z)_\chi = \max_{z\in\mathbb{R}^n}\{z^Tx - \mathbb{1}(x)_\chi\}$

$= \max_{z\in\chi}\{z^Tx\} = S(x)$

~~$S^*(x) = \mathbb{1}(x)_\chi$ [illegible] $(S^*(x))^* = \mathbb{1}(x)_\chi$?~~

~~Since the conjugate of the conjugate~~

$S^*(x) = \mathbb{1}(x)_\chi$ b/c the conjugate of the conjugate is itself for closed, convex, proper functions

b. ~~[illegible] $\max_{z\in\mathbb{R}^n}\{y^Tz - \|x\|\}$~~

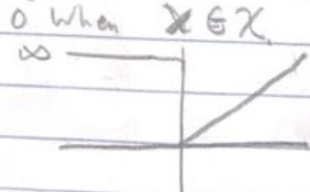
$f(x) = \|x\| = \max_{\|z\|\le 1}\{z^Tx\} = \max_{z\in\chi}\{z^Tx\} = S(x)$

So $f^*(x) = \mathbb{1}(x)$ from part A.

c. $\partial f(x) = \arg\min_{z\in\mathbb{R}^n} f^*(x) = \arg\min_{x\in\mathbb{R}^n} \mathbb{1}(x)_\chi = \chi$

0 when $x \in \chi$.

2 a. $\varphi(x) = \begin{cases} x & x \ge 0 \\ \infty & x < 0 \end{cases}$



a. $S_\lambda(y) = \text{prox}_{\lambda\varphi}(y) = \arg\min\{\frac{1}{2}(x-y)^2 + \lambda\varphi(x)\}$

$$= \begin{cases} \frac{1}{2}(x-y)^2 + \lambda x & x \ge 0 \\ \infty & x < 0 \end{cases}$$

Note no min for ∞ so focus on $x \ge 0$.

Let $x \geq 0$:

$$\operatorname{argmin} \left\{ \frac{1}{2}(x-y)^2 + \lambda x \right\} = x - y + \lambda = 0$$

$x^* = y - \lambda$ for $y > \lambda$ to make x stay > 0.

Let $x = 0$. $\rightarrow$ no ∇ so need g^T for λx.

$$\operatorname{argmin} = x - y + \lambda g^T = 0$$

$$x^* = y - \lambda g^T = 0 \quad y < \lambda$$

$\phi(y) > \phi(x) + g^T(y-x)$

$\frac{\phi(y)}{y} \geq g^T$ since $x=0$.

$\frac{y}{y} = 1 > g^T$

~~$-\infty < g^T < \infty$~~

$-\infty < g^T < 1$

$$\operatorname{prox}_{\lambda g}(y) = \begin{cases} y - \lambda & y > \lambda \\ 0 & y < \lambda. \end{cases}$$

b. $$\operatorname{prox}_{\lambda\|\cdot\|}(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|x-y\|^2 + \lambda\|x\| \right\} = \begin{cases} (\|y\| - \lambda)^+ \frac{y}{\|y\|} & y \neq 0 \\ 0 & y = 0 \end{cases}$$

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \|x\|\|y\| + \lambda\|x\| \right\}$$

$= ?$

3. $x^t = \text{Prox}_{\gamma g}(x^{t-1}) = \text{argmin}\left\{ \frac{1}{2}\|x - x^{t-1}\|^2 + \gamma_t g(x) \right\}$

a. $x^+ = \text{Prox}_{\gamma g}(x) = x - \gamma g_\gamma(x)$ w/ $g_\gamma(x) = \frac{1}{\gamma}(x - \text{Prox}_{\gamma g}(x))$.

$x \in \mathbb{R}^n \quad \gamma > 0 \quad g \in \Gamma^0(\mathbb{R}^n)$

by Theorem 3 of L15 Since $g \in \Gamma^0(\mathbb{R}^n)$ and $z, x \in \mathbb{R}^n$

$x = \text{prox}_g(z) \longleftrightarrow (z - x) \in \partial g(z) \longleftrightarrow (z-x)^T(y-x) \le f(y) - f(x)$

where $z = x^+$

$x^+ - x \in \partial g(\cancel{z}\ x^+) \rightarrow -\gamma g_\gamma(x) \in \partial g(x^+)$

$\text{argmin}\left\{\frac{1}{2}\|x^+ - x\|^2 + \gamma g(x)\right\} = x^+ - x + \gamma \partial g(x) = 0$

$x^+ - x = -\gamma g(x) \rightarrow -\gamma g_\gamma(\cancel{x}) \in -\gamma \partial g(x)$

$\boxed{g_\gamma(x) \in \partial g(x)}$ ✓

b. $g(x^+) \le g_\gamma(\cancel{x}\ z) + \nabla g_\gamma(x)^T(x^+ - \cancel{x}\ z) + \frac{L}{2}\|x^+ - \cancel{x}\ z\|_2^2$

$\gamma \le 1/L$ (prox)

$\le g_\gamma(\cancel{x}\ z) + g_\gamma^T(x^+ - z) - \frac{\gamma}{2}\|g_\gamma(x)\|_2^2$

$= g(\cancel{x}) + g_\gamma(x)^T(x^+ - z) - \frac{\gamma}{2}\|g_\gamma(x)\|_2^2$

$g(x^+) \le g_\gamma(\cancel{x}) + g_\gamma(x)^T(x^+ - \cancel{x}) - \frac{\gamma}{2}\|g_\gamma(x)\|_2^2$

Looking at the inequality we know $\|g_\gamma(x)\|_2^2 > 0$
So subtracting that will give a lower value. Also
$g^T(x)$ is a descent Direction so $g(x^+) \le g(x)$. ✓

c. $\cancel{g \in \Gamma^0(\mathbb{R}^n) \quad g^* = g(x^*) \quad \gamma_t = 1/\sqrt{t}}$

$$\cancel{\frac{\gamma}{2}\sum_{1}^{t} \|g(x^{i-1})\|_2^2 \le g(x^0) - g(x^1) + g(x^1) - g(x^2) + \dots = g(x^0) - g(x^*)}$$

~~where~~ $g(x^*) \le g(x^t)$

$$g(x^t) - g(x^*) \le \frac{1}{2\gamma}\sum\left[\|x^{k-1} - x^*\|_2^2 - \|x^k - x^*\|_2^2\right]$$

$$g(x^t) - g(x^*) \le \frac{1}{2\gamma}\left[\|x^0 - x^*\|_2^2 - \|x^t - x^*\|_2^2\right]$$

$$g(x^t) - g(x^*) \le \frac{1}{2\gamma}\|x^0 - x^*\|_2^2 \qquad \text{Let } \gamma = 1/\sqrt{t}$$

$$g(x^t) - g(x^*) \le \frac{1}{2\sqrt{t}}\|x^0 - x^*\|_2^2$$

d. if $\gamma_t = \gamma > 0$ then $g(x^t) - g(x^*) \ge \frac{1}{2\gamma}\|x^0 - x^*\|_2^2$

?

4. $f(x) = \frac{1}{2}\|Ax-b\|_2^2 \qquad g(x) = \|Ax-b\|_1$

a. $Hf(x) = A^TA \qquad 0 \leq Hf(x) \leq \lambda_{max}(A^TA)I$

So $f(x)$ is weakly convex and smooth w/ Lipschitz constant of ∇f $L = \lambda_{max}(A^TA)$

b. ~~that~~ final value $f =$ ~~4.22~~ 12. I do not believe it is optimal.

c. final value obj $= 3.7 \cdot 10^{-3}$

d. The convergence is not monotone b/c it goes back up a little around the 5th iteration.

e. $g(x) = \|Ax-b\|_1 = \sum_i |Ax_i - b|$

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

$$\partial|x|_1 = \begin{cases} 1 & x>0 \\ -1 & x<0 \\ [-1,1] & x=0 \end{cases}$$

So

$x>0$ or $x<0$ $\nabla g(x) = \pm A$

$x=0 \qquad f(y) \geq f(x) + g^T(y-x)$

$\frac{f(y)}{y} \geq g^T \longrightarrow \frac{|Ay-b|}{y} \geq g^T$

$A - b/y \geq g^T$

$$\partial g(x) = \begin{cases} A^T & x>0 \\ -A^T & x<0 \\ [-A^T, A^T] & x=0 \end{cases}$$

$= |A^T| \, \mathrm{Sgn}(A^Ty - b)$

5. ~~b.~~ a. The final value of g is ~~0.42~~ 0.423.

b. $g(x) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-b_i a_i^T x})$

$\nabla g(x) = \frac{1}{n} A(-b e^{-bA^T x}) / (1 + e^{-bA^T x})$

$Hg(x) = \frac{1}{n} A \,\mathrm{diag}\left[\frac{b^2 \exp(-bA^T x)}{[1+\exp(-bA^T x)]^2} \right] A^T$

$z^T Hg(x) z > 0 \rightarrow$ so $g(x)$ is convex

c. $\partial h(x) = \partial \lambda \|x\|_1 = \lambda \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ [-1,1] & x = 0 \end{cases} = \lambda \,\mathrm{sgn}(x)$

$\mathrm{prox}_{\gamma h}(y) = \mathrm{argmin} \left\{ \frac{1}{2}\|x-y\|^2 + \gamma\lambda \|x\|_1 \right\}$

Let Dimension be 1D.

$x \geq 0 \quad \mathrm{argmin} \left\{ \frac{1}{2}(x-y)^2 + \gamma\lambda x \right\}$

$= (x-y) + \gamma\lambda = 0$

$x = y - \gamma\lambda \qquad y > \gamma\lambda$

$x < 0 \quad \mathrm{argmin} \left\{ \frac{1}{2}(x-y)^2 - \gamma\lambda x \right\}$

$= x - y - \gamma\lambda = 0 \rightarrow x = y + \gamma\lambda$

if $y < -\gamma\lambda$ then $x = y + \gamma\lambda$

if $y > \gamma\lambda$ $x = 0$.

$\mathrm{prox}_{\gamma h}(y) = |y - \gamma\lambda| \,\mathrm{sgn}(y)$