Optimization
Matthew Davidson

Assignment 1

Oct. 6

1. $x^+ = \text{proj}_X(x - \gamma \nabla f(x)) = x - \gamma g(x)$

a.
$$\nabla f(x)^T (z - x^*) \geq 0$$
$$-(x^* - y)^T (z - x^*) \geq 0$$
$$(y - x^*)^T (z - x^*) \leq 0$$
$$\boxed{(y - \text{proj}_X(y))^T (z - \text{proj}_X(y)) \leq 0}$$

b. Let $y = x - \gamma \nabla f(x)$

Then: $(y - \text{proj}_X(y))^T (z - \text{proj}_X(y)) \leq 0$
$$(x - \gamma \nabla f(x) - \text{proj}_X(x - \gamma \nabla f(x)))^T (z - x^+) \leq 0$$
$$(\gamma g(x) - \gamma \nabla f(x))^T (z - x^+) \leq 0$$
$$\gamma g(x)^T (z - x^+) - \gamma \nabla f(x)^T (z - x^+) \leq 0$$
$$\boxed{g(x)^T (x^+ - z) \geq \nabla f(x)^T (x^+ - z)}$$

c.
$$f(x^+) \leq f(x) + \nabla f(x)^T (x^+ - x) + \frac{L}{2} \|x^+ - x\|_2^2$$
$$\leq f(x) + \nabla f(x)^T (-\gamma g(x)) + \frac{L}{2} \|-\gamma g(x)\|_2^2$$
$$= f(x) - \gamma \nabla f(x)^T g(x) + \frac{L \gamma^2}{2} \|g(x)\|_2^2$$

note: $\gamma = 1/L$

$$= f(x) - \gamma \nabla f(x)^T g(x) + \frac{\gamma}{2} \|g(x)\|_2^2$$

note: 1b.

$$\leq f(x) - \gamma g(x)^T g(x) + \frac{\gamma}{2} \|g(x)\|_2^2$$
$$= f(x) - \gamma \|g(x)\|_2^2 + \frac{\gamma}{2} \|g(x)\|_2^2$$

$$\boxed{\therefore f(x^+) \leq f(x) - \frac{\gamma}{2} \|g(x)\|_2^2}$$

1d. $f(x^{t+1}) \leq f(x^{0}) - \frac{\gamma}{2}\|g(x)\|_2^2 \leq f(x^*) + \nabla g(x)^T(x-x^*) - \frac{\gamma}{2}\|g(x)\|_2^2$

$= f(x^*) + \frac{1}{2\gamma}\left[\|x-x^*\|_2^2 - \|x-x^*\|_2^2 + 2\gamma g(x)^T(x-x^*) - \gamma^2\|g(x)\|_2^2\right]$

$= f(x^*) + \frac{1}{2\gamma}\left[\|x-x^*\|_2^2 - \|x-x^*-\gamma g(x)\|_2^2\right]$

$= f(x^*) + \frac{1}{2\gamma}\left[\|x-x^*\|_2^2 - \|x^+-x^*\|_2^2\right]$

now consider $x = x^{k-1}$ and $x^+ = x^k$ $\therefore$ $k = 1 \dots t$

$$\sum_{k=1}^{t}\left(f(x^k) - f(x^*)\right) \leq \frac{1}{2\gamma}\sum\left[\|x^{k-1}-x^*\|_2^2 - \|x^k - x^*\|_2^2\right]$$

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

$$\leq \frac{1}{2\gamma}\|x^0 - x^*\|_2^2$$

So $f(x^t) - f(x^*) \leq -\frac{\gamma}{2}\|g(x)\|_2^2 \leq f(x^t) - f(x^*) \leq$

$$\leq \frac{1}{t}\sum_{k=1}^{t}\left(f(x^k) - f(x^*)\right) \leq \frac{1}{2\gamma t}\|x^0 - x^*\|_2^2$$

$\therefore$ as $t \to \infty$, $g(x^t) \to 0$

7. $|x^Ty| \le \|x\| \|y\|_*$ where $\|y\|_* = \max_{\|x\|\le 1} x^Ty$

a. $x^Ty \le \frac{x^Ty}{\|x\|} \|x\| \le \|x\| \max_{\|x\|\le 1} \frac{x^Ty}{\|x\|} = \|x\| \|y\|_*$

$x^Ty \le \|x\| \|y\|_*$

$-x^Ty \ge -\frac{x^Ty}{\|x\|} \|x\| \ge -\|x\| \max_{\|x\|\le 1} \frac{x^Ty}{\|x\|} = -\|x\| \|y\|_*$

$\therefore \boxed{|x^Ty| \le \|x\| \|y\|_*}$.

b. $f(x) = \|x\|_\infty = \max\{|x_1| \dots |x_n|\}$

Let $x \neq 0$ $\quad \partial \|x\|_\infty = \bigcap_{y \in \mathbb{R}^n} H_y$ where $H_y = \{g \in \mathbb{R}^n : f(y) \ge f(x) + g^T(y-x)\}$

$\hookrightarrow$ active functions

$\partial \|x\|_\infty = \cap H_y = \mathrm{conv}\left(\bigcup \{\partial f_i(x) : f_i(x) = f(x)\} \right)$

$\partial \|x\|_\infty = \bigcup_{i=1}^n \partial f_i(x) = \begin{cases} -1 & x \le 0 \\ 1 & x > 0 \end{cases} = \mathrm{sgn}(x_i) e_i$

Let $x = 0$ $\quad g \in \partial f(x) \longleftrightarrow \|y\|_\infty = f(y) \ge f(0) + g^T(y-0) = g^Ty$

$g^Ty \le \|g\|_1 \|y\|_\infty \le \|y\|_\infty$

$\partial f(x) = \begin{cases} \mathrm{sgn}(x_i) & x_i \neq 0 \\ \{g \in \mathbb{R}^n : \|g\|_1 \le 1 \end{cases}$

3. $f(x) = \|x\|$ $\quad Z(x) := \{z \in \mathbb{R}^n : z^T x = \|x\|, \|z\|_* \le 1\}$

a. Suppose $\forall z \in Z(x)$ and $y \in \mathbb{R}^n$

$$f(y) \ge f(x) + g^T(y-x) \longrightarrow \|x\| + z^T(y-x) = \|x\| + z^T y - z^T x$$

Since $z \in Z(x) \to z^T x = \|x\|$ So,

$$= \|x\| + z^T y - \|x\| = z^T y \le \|y\|\,\|z\|_* \overset{\|z\|_* \le 1}{\le} \|y\| = f(y).$$

$\therefore Z(x) \in g^T \to g^T \in \partial f(x)$ So $Z(x) \in \partial f(x)$

b. Consider $z \in \partial f(x)$.

$$\|y\| = f(y) \ge \|x\| + z^T(y-x)$$
$$z^T y - \|y\| \le z^T x - \|x\|$$
$$\max\{z^T y - \|y\|\} \le z^T x - \|x\|$$

$$\left.\begin{matrix} \|z\|_* \le 1 & 0 \\ \text{else} & \infty \end{matrix}\right\} = \mathbb{1}_{\|\cdot\|_* \le 1}(z) \le z^T x - \|x\|$$

$$0 \le z^T x - \|x\| \le \|x\|\,\|z\|_* - \|x\| = \|x\|(\|z\|_* - 1)$$

Since $\|z\|_* \le 1$ Then

$$0 \le \|x\|(\|z\|_* - 1) \le 0$$
So $0 \le z^T x - \|x\| \le 0$

$\therefore z^T x = \|x\|$ So $\partial f(x) \subseteq Z(x)$

c. ~~$\|x$~~ max

$$\|x\|\,\|z\|_* = z^T x \longrightarrow \|x\| = \frac{z^T x}{\|z\|_*} = \max_{\|z\|_* \le 1} z^T x$$

$$\boxed{\partial\|x\| = \arg\max\{z^T x\}}$$

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

a. $\nabla f(x) = A^T(Ax-b) = A^TAx - A^Tb$

$Hf(x) = A^TA = \|A\|_2^2$

Since $\|A\|_2^2$ is always positive Then f is convex.

f is smooth b/c is has infinitely many derivatives, Just after $Hf(x)$ Then it is Just 0.

$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\,\|x-y\|_2$

$$L \geq \frac{\|\nabla f(x) - \nabla f(y)\|_2}{\|x-y\|_2} = \frac{\|A^TAx - A^Tb - A^TAy + A^Tb\|_2}{\|x-y\|_2}$$

$$= \frac{\|A^TAx - A^TAy\|_2}{\|x-y\|_2} = \frac{A^TA\,\|x-y\|_2}{\|x-y\|_2} = A^TA$$

$L \geq \|A\|_2^2$

b. Submitted as attachment.

c. final error = 0.767 it can't compute The True value because The The prediction for The True value's were not restricted To The ~~Domain~~ Same domain as The True values (b).

So although it got closer To The objective, it wasn't a good predictor. If minimizing over R^n_+ Then all The negative predicted values would be 0 and decrease The error.

d. See attachment.

e. The GM got closer To The objective but The PGM had less error because it restricted The domain To nonnegative values.

5. (P) minimize $f(x) = \frac{1}{2}\|Ax-b\|_2^2 \quad \|x\|_1 \le \tau$

$\tau > 0$, $A \in \mathbb{R}^{m \times n}$ $b \in \mathbb{R}^m$

a. for $x \neq 0$ $i \in \{1, \dots, n\}$ w/ $|x_i| = \max_j |x_j|$

from Problem 2

$$\partial f(x) = \operatorname{conv}\left(\cup\{\partial f_i(x) : f_i(x) = f(x)\}\right) = \operatorname{sgn}(x_i) e_i$$

So $\operatorname{sgn}(x_i) e_i \in \partial \|x\|_\infty$

b. from problem 3

$$\partial \|\nabla f(x^{t-1})\|_\infty = \underset{\|z\|_1 \le 1}{\operatorname{argmax}} \{\nabla f(x^{t-1})^T z\}$$

$$-\partial \|\nabla f(x^{t-1})\|_\infty = \underset{\|z\|_1 \le 1}{\operatorname{argmin}} \{\nabla f(x^{t-1})^T z\}$$

$$-\tau \partial \|\nabla f(x^{t-1})\|_\infty = \underset{\|z\|_1 \le 1}{\operatorname{argmin}} \{\nabla f(x^{t-1})^T \tau z\}$$

Let $s = \tau z$

$$\boxed{-\tau \partial \|\nabla f(x^{t-1})\|_\infty = \underset{\|s\|_1 \le \tau}{\operatorname{argmin}} \{\nabla f(x^{t-1})^T s\}}$$

c. $s^t = \underset{\|s\|_1 \le \tau}{\operatorname{argmin}} \{\nabla f(x^{t-1})^T s\} = -\tau \partial \|\nabla f(x^{t-1})\|_\infty = -\tau \operatorname{sgn}(x_i) e_i$

$s^t = -\tau \operatorname{sgn}(x_i) e_i$ where $i \in \underset{i \in (1, \dots, n)}{\operatorname{argmax}} \{\nabla f(x^{t-1})_i\}$

$x^t = x^{t-1} + \gamma_t (s^t - x^{t-1}) = x^{t-1} + \gamma_t s^t - \gamma_t x^{t-1} = (1-\gamma_t) x^{t-1} + \gamma_t s^t$

So

$x^t = (1-\gamma_t) x^{t-1} + \gamma_t s^t.$

$m=32$
$n=64$
$s=n$

$A=32\times 64$ $\quad s=64\times 1$ $\quad b=32\times 1$
$x=64\times 1$

5d. $\underset{\gamma\in[0,1]}{\arg\min}\ f((1-\gamma)x^{t-1}+\gamma s^t) = \underset{\gamma}{\arg\min}\ \frac{1}{2}\|A[(1-\gamma)x^{t-1}+\gamma s^t]-b\|_2^2$

$$=\underset{\gamma}{\arg\min}\ \frac{1}{2}\|Ax^{t-1}-\gamma Ax^{t-1}+\gamma As^t-b\|_2^2$$

$$=(Ax-\gamma Ax+\gamma As-b)(As-Ax)=0$$

$$(Ax)^T(As-(Ax)^TAx-\gamma(Ax)(As)^T+\gamma(Ax)^TAx+\gamma(As)^TAs$$
$$-\gamma(As)^TAx-bAs+bAx=0$$

$$\gamma(A^TS^TAS-Ax(AS)^T-A^TS^TAx+A^Tx^TAx)$$
$$+x^TA^Tb-S^TA^Tb+S^TA^TAx-x^TA^TAx=0$$

$$\gamma(A^TS^TAS-A^TS^TAx+A^TS^TAS+A^Tx^TAx)=S^TA^Tb-x^TA^Tb-S^TA^TAx+x^TA^TAx$$

$$\gamma\|A(s^t-x^{t-1})\|_2^2=(s^t-x^{t-1})^TA^T(b-Ax^{t-1})$$

$$\gamma=\frac{(s^t-x^{t-1})^TA^T(b-Ax^{t-1})}{\|A(s^t-x^{t-1})\|_2^2}$$

since $\gamma\in[0,1]$

$$\boxed{\gamma_t=\text{proj}_{[0,1]}\left(\frac{(s^t-x^{t-1})^TA^T(b-Ax^{t-1})}{\|A(s^t-x^{t-1})\|_2^2}\right)}$$

note: $\underbrace{A^TS^TAS-A^TS^TAx+A^TS^TAS+A^Tx^TAx}_{\text{factorable}}=(As^t-Ax^{t-1})^T(As^t-Ax^{t-1})$

also: $(s^t-x^{t-1})^TA^T(b-Ax^{t-1})=(S^TA^Tb-x^{t-1T}A^Tb-S^TA^TAx^{t-1}+x^{t-1T}A^TAx^{t-1})$

e. The error for the CGm is 0.179 which is lower than the Pgm, 0.23. This is because we keep the step size in the CGm between [0,1].