Homework 5

Yujun Kim

IE539 Convex Optimization

Dec 2023

1 Moreau-Yosida Smoothing of 1-norm

(Step 0) Result from Homework 2. From homework2 problem 1-(b), we found the closed form solution of the unconstrained problem

$$f(x) = \frac{1}{2}||x - z||_2^2 + \lambda||x||_1$$

as following equation

$$\begin{aligned} x_i^* &= \begin{cases} z_i - \lambda & \text{if } z_i > \lambda \\ 0 & \text{if } -\lambda \leq z_i \leq \lambda \\ z_i + \lambda & \text{otherwise} \end{cases} \\ &= max\{z_i - \lambda, 0\} + min\{z_i + \lambda, 0\} \end{aligned}$$

For completeness, I copied my answer from homework 2 to section 3(Appendix below).

(Step 1). Let $g(x, u) = \frac{1}{\eta}(\frac{1}{2}||u - x||^2 + \eta||u||_1)$. Then,

$$\begin{aligned} f_\eta(x) &= \inf_u\{f(u) + \frac{1}{2\eta}||u - x||^2\} \\ &= \frac{1}{\eta}\inf_u\{\frac{1}{2}||u - x||^2 + \eta||u||_1\} \\ &= g(x, u^*) \end{aligned}$$

where the last inequality holds by step 0 with

$$u_i^* = \begin{cases} x_i - \eta & \text{if } x_i > \eta \\ 0 & \text{if } -\eta \leq x_i \leq \eta \\ x_i + \eta & \text{otherwise} \end{cases}$$

We expand the square norm as sum of squares of element in each dimension.

$$\begin{aligned} f_\eta(x) &= g(x, u^*) \\ &= \frac{1}{\eta}(\frac{1}{2}||u^* - x||^2 + \eta||u^*||_1) \\ &= \sum_{i=1}^d \frac{1}{\eta}\left(\frac{1}{2}(u_i^* - x_i)^2 + \eta|u_i^*|\right) \end{aligned}$$

Hence, it only remains to prove

$$L_\eta(x_i) = \frac{1}{2}(u_i^* - x_i)^2 + \eta|u_i^*|$$

(Step 2). We prove equality above, by dividing cases into three.

- If $x_i > \mu$, then $u_i^* = x_i - \eta$. Thus,

$$\begin{aligned}\frac{1}{2}(u_i^* - x_i)^2 + \eta|u_i^*| &= \frac{1}{2}(-\eta)^2 + \eta|x_i - \eta| \\ &= \frac{1}{2}\eta^2 + \eta(x_i - \eta) \\ &= \eta x_i - \frac{1}{2}\eta^2 \\ &= \eta|x_i| - \frac{1}{2}\eta^2 \\ &= L_\eta(x_i)\end{aligned}$$

- If $\mu \geq x_i \geq -\mu$, then $u_i^* = 0$. Thus,

$$\begin{aligned}\frac{1}{2}(u_i^* - x_i)^2 + \eta|u_i^*| &= \frac{1}{2}(x_i)^2 \\ &= L_\eta(x_i)\end{aligned}$$

- If $x_i < -\mu$, then $u_i^* = x_i + \eta$. Thus,

$$\begin{aligned}\frac{1}{2}(u_i^* - x_i)^2 + \eta|u_i^*| &= \frac{1}{2}\eta^2 + \eta|x_i + \eta| \\ &= \frac{1}{2}\eta^2 - \eta(x_i + \eta) \\ &= -\eta x_i - \frac{1}{2}\eta^2 \\ &= \eta|x_i| - \frac{1}{2}\eta^2 \\ &= L_\eta(x_i)\end{aligned}$$

2 Dual Subgradient/ Proximal Point update of Dual

(a) Dual Gradient Method Observe the following equivalence.

$$\begin{aligned}&g_t \in \partial h(\mu_t) = \partial(f^*(-A^T\mu_t) + g^*(\mu_t)) = -A\partial f^*(-A^T\mu_t) + \partial g^*(\mu_t) \\ &\quad\Leftrightarrow g_t = -Ax_t + y_t \text{ for some } x_t \in \partial f^*(-A^T\mu_t), \quad y_t \in \partial g^*(\mu_t) \\ &\quad\Leftrightarrow g_t = -Ax_t + y_t \text{ with } -A^T\mu_t \in \partial f(x_t), \quad \mu_t \in \partial g(y_t) \\ &\quad\Leftrightarrow g_t = -Ax_t + y_t \text{ for some } x_t \in argmin_x\{f(x) + \mu_t^T Ax\}, \quad y_t \in argmin_y\{g(y) - \mu_t^T y\}\end{aligned}$$

Thus, $\mu_{t+1} = \mu_t - \eta_t g_t$ for some $g_t \in \partial h(\mu_t)$ if and only if

$$\mu_{t+1} = \mu_t + \eta(Ax_t - y_t)$$

for some

$$x_t \in argmin_x\{f(x) + \mu_t^T Ax\}, \quad y_t \in argmin_y\{g(y) - \mu_t^T y\}$$

(b)Proximal point method on dual is Augmented Lagrangian Method. Observe the following equivalence.

$$\begin{aligned}
&\mu_{t+1} = prox_{\eta h}(\mu_t) \\
&\quad \Leftrightarrow \mu_{t+1} = argmin_\mu\{f^*(-A^T\mu) + g^*(\mu) + \frac{1}{2\eta}||\mu - \mu_t||^2\} \\
&\quad \Leftrightarrow 0 \in -A\partial f^*(-A^T\mu_{t+1}) + \partial g^*(\mu_{t+1}) + \frac{1}{\eta}(\mu_{t+1} - \mu_t) \\
&\quad \Leftrightarrow \mu_{t+1} = \mu_t + \eta(Ax_t - y_t) \text{ for some } x_t \in \partial f^*(-A^T\mu_{t+1}), \quad y_t \in \partial g^*(\mu_{t+1}) \\
&\quad \Leftrightarrow \mu_{t+1} = \mu_t + \eta(Ax_t - y_t) \text{ for some } -A^T\mu_{t+1} \in \partial f(x_t), \quad \mu_{t+1} \in \partial g(y_t) \\
&\quad \Leftrightarrow \mu_{t+1} = \mu_t + \eta(Ax_t - y_t) \text{ for some} \\
&\qquad 0 \in \partial f(x_t) + A^T(\mu_t + \eta(Ax_t - y_t)), \quad 0 \in \partial g(y_t) + \mu_t + \eta(Ax_t = y_t)) \\
&\quad \Leftrightarrow \mu_{t+1} = \mu_t + \eta(Ax_t - y_t) \text{ for some} \\
&\qquad (x_t, y_t) \in argmin\{f(x) + g(y) + \mu_t^T(Ax - y) + \frac{\eta}{2}||Ax - y||^2\}
\end{aligned}$$

Thus, the statement holds.

(c) h is convex for being a sum of convex functions. Proposition 20.7 of lecture note states that

$$\nabla h_\eta(x) = prox_{h^*/\eta}\left(\frac{x}{\eta}\right) = \frac{1}{\eta}(x - prox_{\eta h}(x))$$

Substitute $x = \mu_t$ to obtain, $prox_{\eta h}(\mu_t) = \mu_t - \eta \nabla h_\eta(\mu_t)$.

3 Appendix for Problem 1

Let $f(x) = \frac{1}{2}||x - z||_2^2 + \lambda||x||_1$ be the objective function. Then for $x_i \neq 0$,

$$\frac{\partial f}{\partial x_i} = x_i - z_i + \lambda \cdot \text{sgn}(x_i)$$

where $\text{sgn}(t)$ is the sign of t. We divide cases into three: $z_i > \lambda, z_i < -\lambda, -\lambda \leq z_i \leq \lambda$. For each case, we analyze by considering both $x_i > 0$ and $x_i < 0$. Note that f being not differentiable at $x_i = 0$ does not matter by the continuity of f.

- If $z_i > \lambda$. By considering both $x_i > 0, x_i < 0$, for $x_i \neq 0$

  $$\frac{\partial f}{\partial x_i}\begin{cases} > 0 \text{ if } x_i > z_i - \lambda \\ < 0 \text{ if } x_i < z_i - \lambda \end{cases}$$

  Hence, x_i minimize f when $x_i = z_i - \lambda$.

- If $z_i < -\lambda$. By considering both $x_i > 0, x_i < 0$, for $x_i \neq 0$

  $$\frac{\partial f}{\partial x_i}\begin{cases} > 0 \text{ if } x_i > z_i + \lambda \\ < 0 \text{ if } x_i < z_i + \lambda \end{cases}$$

  Hence, x_i minimize f when $x_i = z_i + \lambda$

- If $-\lambda \leq z_i \leq \lambda$. By considering both $x_i > 0, x_i < 0$, for $x_i \neq 0$

$$\frac{\partial f}{\partial x_i} \begin{cases} > 0 \text{ if } x_i > 0 \\ < 0 \text{ if } x_i < 0 \end{cases}$$

  Hence, x_i minimize f when $x_i = 0$.

Thus, we have closed for for x^*.

$$\begin{aligned} x_i^* &= \begin{cases} z_i - \lambda & \text{if } z_i > \lambda \\ 0 & \text{if } -\lambda \leq z_i \leq \lambda \\ z_i + \lambda & \text{otherwise} \end{cases} \\ &= max\{z_i - \lambda, 0\} + min\{z_i + \lambda, 0\} \end{aligned}$$

The optimal value of the objective function is $f(x^*)$.