Hw3 - a

To prove $Aw-b \in \partial g(\lambda)$, we should show that $Aw-b$ satisfy the $\partial g(\lambda)$ definition,

$$g(\lambda_1) \geq g(\lambda_2) + \partial g(\lambda_2)(\lambda_1 - \lambda_2)$$

$$\hookrightarrow \cancel{f(w)} + \lambda_1^T (Aw-b) \geq \cancel{f(w)} + \lambda_2^T (Aw-b) + \partial g(\lambda_2)^T [\lambda_1 - \lambda_2]$$

$$\Rightarrow (\lambda_1 - \lambda_2)^T (Aw-b) \geq \partial g(\lambda_2)^T [\lambda_1 - \lambda_2]$$

Now, we check if $Aw-b$ satisfies this condition $\longrightarrow (\lambda_1 - \lambda_2)^T [Aw-b] \overset{?}{\geq} [Aw-b] [\lambda_1 - \lambda_2]$

$\rightsquigarrow$ This inequality is true & actually the equality holds, therefore ∂g can be equal to $Aw-b$, which means

$Aw-b \in \partial g(\lambda)$

~~Thane~~

HW 3_b

There are 2 theorems as below:

① If f is closed & strong convex with parameter μ, then f^* has a Lipschitz continuous gradient with parameter μ.

② If f is convex and has a Lipschitz continuous with parameter L, then f^* is strong convex with parameter $1/L$.

proof of (1): By implication of strong convexity, we have $\|s_x - s_y\| \geq \mu \|x-y\|$ $\forall s_x \in \partial f(x), s_y \in \partial f(y)$

which implies $\|s_x - s_y\| \geq \mu \|\nabla f^*(s_x) - \nabla f^*(s_y)\|$

Hence, f^* has a Lipschitz continuous ~~with~~ gradient with $1/\mu$

proof of (2): By implication of Lipschitz continuous gradient for convex f, we have:

continue of HW 3 - b →

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2$$

which implies:

$$(s_x - s_y)^T (x-y) \geq \frac{1}{L} (s_x - s_y) \quad \forall x \in \partial f^*(s_x), \; y \in \partial f^*(s_y)$$

Hence, f^* is strongly convex with parameter $1/L$. $\square$

Therefore, the convergence rate is same as primal gradient descent problem with

$$\alpha_k = \frac{2}{\frac{1}{\mu} + \frac{1}{L}}$$

as step size, and of course with linear rate.

→ primal feasible?

The solution is feasible, because the dual problem converges. Accordingly, the primal solution would be ~~feasib~~ feasible.

HW 3_C

$$\text{minimize } \frac{1}{N} \sum_{i \in [N]} f_i(w_i)$$

$$\text{s.t } w_i = w_j \text{ for all } j \in N_i$$

$$L(w,\lambda) = \frac{1}{N} \sum_i \left(f_i(w_i) - \sum_{j \in N_i} \lambda_{ij} (w_i - w_j) \right)$$

$$L(w,\lambda) = \frac{1}{N} \sum_i \left(f_i(w_i) - \sum_{j \in N_i} \lambda_{ij} w_i + \sum_{j \in N_i} w_j \right.$$

Step ① node i computes :

$$w_i^k = \arg\min_w \frac{1}{N} f_i(w) - \sum_{j \in N_i} \lambda_{ij} w_i$$

Step ② : node i sends w_i to N_i neighbors

Communication cost = number of nodes x average node degree.

Then compute :

$$\lambda_{ij}^{k+1} = \lambda_{ij}^{k} + \alpha^k (w_i - w_j) \text{ for all } j \in N_i$$

Convergence rate of the dual former is lower than primal because it is based on one hop communication.