HW3

3.1) f is convex and closed, so we can find ω^* where

$$\omega^* = \underset{\omega}{\text{argmin}}\ f(\omega) + \lambda^T(A\omega - b)$$

$$g(\lambda) = \inf_{\omega} L(\omega,\lambda) = f(\omega^*) + \lambda^T(A\omega^* - b)$$

$\forall \mu$ $\quad g(\mu) = \inf_{\omega} L(\omega,\mu) \leq L(\omega^*,\mu) = f(\omega^*) + \mu^T(A\omega^* - b)$

$$\Rightarrow g(\mu) \leq g(\lambda) + (\mu - \lambda)^T(A\omega^* - b) \Rightarrow A\omega^* - b \in \partial g(\lambda) \quad \forall \mu$$

3.2) if f is μ-strongly convex, dual ascent algorithm is as follows:

$$\begin{cases} \omega_{k+1} = \underset{\omega}{\text{argmin}}\ L(\omega, \lambda_k) \\ \lambda_{k+1} = \lambda_k + \alpha_k(A\omega_k - b) \end{cases}$$

Note that $g(\lambda) = \inf_{\omega} L(\omega,\lambda) = f(\omega^*) + \lambda^T(A\omega^* - b) \Rightarrow \nabla g(\lambda) = A\omega^* - b$

So $\lambda_{k+1} = \lambda_k + \alpha_k \nabla g(\lambda_k)$ .. Then we can use results for GD and convergence for smooth and strongly convex function.

$f(\cdot)$ is μ-strongly convex and L-smooth. Then $g(\lambda) = \inf_{\omega} f(\omega) + \lambda^T(A\omega - b)$ is $\frac{1}{L}$ strongly convex and $\frac{1}{\mu}$ smooth (Lecture 2 - food for thought)

From convergence rate of GD (Lecture 2 - Theorem 1) we have

$$\|\lambda_k - \lambda^*\|^2 \leq \left(1 - \frac{2}{1 + \mu/L}\right)^{2k} \|\lambda_0 - \lambda^*\|^2 = \left(\frac{\mu - L}{L + \mu}\right)^{2k} \|\lambda_0 - \lambda^*\|^2 \quad \text{for } \alpha = \frac{2\mu L}{\mu + L}.$$

Is solution primal feasible?

Assume exact convergence. $\omega_{k+1} = \omega_k$ for some k

$$\underset{\omega}{\text{argmin}}\ L(\omega, \lambda_{k+1}) = \underset{\omega}{\text{argmin}}\ L(\omega, \lambda_k) \qquad \lambda_{k+1} = \lambda_k + \alpha_k(A\omega_k - b)$$

$$f(\omega) + (\lambda_k + \alpha_k(A\omega_k - b))^T(A\omega - b) = f(\omega) + \lambda_k(A\omega - b)$$

$$\Rightarrow \alpha_k \|A\omega_k - b\|^2 = 0 \Rightarrow A\omega_k = b$$

Alternatively at convergence $\nabla g(\lambda_k) = 0 \Rightarrow A\omega_k = b$

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

s.t.

$w_i = w_j$

$\forall j \in N_i$

Define $L(w_1, \dots, w_N, \lambda_{11}, \dots \lambda_{NN}) = \frac{1}{N} \sum_{i \in [N]} f_i(w_i) + \sum_{i \in [N]} \sum_{j \in N_i} \lambda_{ij}^T (w_i - w_j)$

$$= \frac{1}{N} \sum_{i=1}^{N} f_i(w_i) + \sum_{i=1}^{N} a_i^T w_i$$

where $a_i = \sum_{j \in N_i} \lambda_{ij} - \lambda_{ji}$. so we have

$$g(\lambda_{11}, \dots, \lambda_{NN}) = \min_{w_1, \dots w_N} L(w_1, \dots, w_N, \lambda_{11}, \dots, \lambda_{NN})$$

$$= \min_{w_1 \dots w_N} \sum_{i=1}^{N} \frac{1}{N} f_i(w_i) + a_i^T w_i$$

$$\nabla g(\lambda_{11} \dots \lambda_{NN}) = \begin{cases} w_i^* - w_j^* & j \in N_i \\ 0 & \text{o.w.} \end{cases}$$

$$\begin{cases} 1) \; w_{i,k+1} \in \operatorname{argmin}_{w_i} \frac{1}{N} f_i(w_i) + a_i^T w_i \\ 2) \; \lambda_{ij,k+1} = \lambda_{ij} + \alpha_k (w_{i,k+1} - w_{j,k+1}) \quad j \in N_i \end{cases}$$