Fundamentals of Machine Learning Over Networks

Group 6
HW 3

HW3(a):

1) $g(\lambda)$ is concave

$$g(t\lambda_1+(1-t)\lambda_2) = \inf_{w} f(w) + (t\lambda_1+(1-t)\lambda_2)^T(Aw-b)$$

$$= \inf_{w} t[f(w)+\lambda_1^T(Aw-b)] + (1-t)[f(w)+\lambda_2^T(Aw-b)]$$

$$\geqslant t \inf_{w} f(w)+\lambda_1^T(Aw-b) + (1-t)\inf_{w} f(w)+\lambda_2^T(Aw-b)$$

$$= t\,g(\lambda_1) + (1-t)\,g(\lambda_2) \Rightarrow \checkmark$$

2) Assume that the domain of f is convex and compact $\rightarrow$ $L(w,\lambda)$ has a minimum. So,

$w^* \in \underset{w}{\operatorname{argmin}}\ f(w) + \lambda^T(Aw-b)$

We must show that $g(\mu) \leqslant g(\lambda) + (Aw^*-b)^T(\mu-\lambda) \quad \forall \mu$

$g(\lambda) = f(w^*) + \lambda^T(Aw^*-b) \rightarrow g(\lambda) + (Aw^*-b)^T(\mu-\lambda) = f(w^*) + (Aw^*-b)\mu \geqslant g(\mu) \rightarrow \checkmark$

HW3 (b):

$*$ If $f(x)$ is μ-strongly convex and L-smooth over a convex and closed set, then $f^*(y) = \sup_x \; y \cdot x - f(x)$ is $\frac{1}{L}$-strongly convex and $\frac{1}{\mu}$-smooth.

In dual ascent we have

$$\begin{cases} w_{k+1} \in \operatorname{argmin}_w L(w, \lambda_k) \\ \lambda_{k+1} = \lambda_k + \alpha_k (A w_k - b) \end{cases} \Longrightarrow \lambda_{k+1} = \lambda_k + \alpha_k \overbrace{\partial g(\lambda_k)}^{\text{from (a)}} \longrightarrow \text{gradient descent to find } \max_\lambda g(\lambda)$$

So, for fixed $\alpha = \dfrac{2}{\frac{1}{\mu} + \frac{1}{L}} = \dfrac{2L\mu}{L+\mu}$ we have

$$\|\lambda_k - \lambda^*\|^2 \leqslant \left(\frac{\mu - L}{L+\mu}\right)^{2k} \|x_0 - x^*\|^2 \longleftarrow \text{since } g(\lambda) \text{ is } \tfrac{1}{L}\text{-strongly concave and } \tfrac{1}{\mu} \text{ smooth}$$

($*$) was proved in "foods for thought" in Lecture 2. (Boyd, Problem 3.40)

HW3(c):

By writing $L(w_1,\dots,w_N,\lambda_{11},\dots,\lambda_{NN})$

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

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

where $\boxed{a_i = \sum_{j\in N_i}\lambda_{ij}-\lambda_{ji}}$. Hence,

$$g(\lambda_{11},\dots,\lambda_{NN}) \triangleq \min_{w_1,\dots,w_N} L(w_1,\dots,w_N,\lambda_{11},\dots,\lambda_{NN}) = \sum_{i=1}^{N} g_i(a_i),$$

where $g_i(a_i) \triangleq \min_{w_i}\frac{1}{N}f_i(w_i)+a_i^T w_i$

Likewise part (a), we can prove that $w_i^* \in \partial g_i(a_i)$

$$\frac{\partial}{\partial\lambda_{ij}} g(\{\lambda_{ij}\}) = \sum_{k=1}^{N} w_k^{*T}\frac{\partial}{\partial\lambda_{ij}}a_k = (w_i^{*T}-w_j^{*T})\,\mathbb{1}(j\in N_i) \Rightarrow \nabla_{ij} g = \begin{cases} w_i^*-w_j^* & j\in N_i \\ 0 & \text{o.w.}\end{cases}$$

1) $w_{i,k+1} \in \arg\min_{w_i} \frac{1}{N}f_i(w_i)+a_i^T w_i$

2) $\lambda_{ij,k+1} = \lambda_{ij,k} + \alpha_k(w_{i,k+1}-w_{j,k+1})$