HW3 (a). $\min_x f(x)$ s.t. $Ax=b$

For convex and closed function f, there is $f^{**}=f$

$x \in \partial f^*(y) \Rightarrow y \in \partial f^{**}(x) \Rightarrow y \in \partial f(x)$.

$y \in \partial f(x) \Rightarrow f(u) - f(x) \geq y^T(u-x), \quad \forall u$.

$\Rightarrow y^T u - f(u) \leq y^T x - f(x), \quad \forall u$.

$\Rightarrow f^*(y) = \max_u (y^T u - f(u)) = y^T x - f(x) \Leftrightarrow x = \arg\min_u f(u) - y^T u$

also let $f^*(v) = \max_u \{v^T u - f(v)\} \geq f^*(y) + x^T(v-y) \Rightarrow x \in \partial^*(y)$

The dual problem: $\max_\lambda g(\lambda) = -f^*(-A^T\lambda) - \lambda^T b$

$$\frac{\partial g(\lambda)}{\partial \lambda} = \frac{A \partial f^*(-A^T\lambda)}{\partial \lambda} - b$$

Since for convex and close function f

$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x = \arg\min_u f(u) - y^T u$.

$\Rightarrow w \in \partial f^*(-A^T\lambda) \Leftrightarrow w = \arg\min_u f(u) + \lambda^T A w$.

$\Rightarrow Aw - b \in \partial g(\lambda) \Leftrightarrow w = \arg\min_w f(w) + \lambda^T A w$.

Thus, $Aw - b \in \partial g(\lambda)$ for $w = \arg\min_u f(u) + \lambda^T A w$.

HW3(b) $\min f(w)$

$\text{s.t. } Aw = b$

since $w_{k+1} \in \arg\min_{w} L(w, \lambda_k)$, where

$L(w, \lambda_k) = f(w) + \lambda_k^T (Aw - b)$

Denote the primal optimal variable by w^*

$$L(w^*, \lambda_k) - L(w_{k+1}, \lambda_k)$$
$$= f(w^*, \lambda_k) - f(w_{k+1}, \lambda_k) + \lambda_k^T (Aw^* - b) - \lambda_k^T (Aw^* - b)$$
$$\geq \nabla f(w_{k+1})^T (w^* - w_{k+1}) + \frac{\mu}{2} \|w^* - w_{k+1}\| + \lambda_k^T A (w^* - w_{k+1})$$

where the above inequality follows from f is μ-strongly convex.

since $w_{k+1} \in \arg\min_{w} L(w, \lambda_k)$

$$\nabla L(w_{k+1}, \lambda_k) = 0 \Rightarrow f(w_{k+1}) + \lambda_k^T A = 0$$

Thus we have

$$L(w^*, \lambda_k) - L(w_{k+1}, \lambda_k) \geq \frac{\mu}{2} \|w^* - w_{k+1}\|^2$$

in another word

$$\|w^* - w_{k+1}\|^2 \leq \frac{2(L(w^*, \lambda_k) - L(w_{k+1}, \lambda_k))}{\mu}.$$

It can be seen that the convergence and accuracy of primal can be controlled by dual variable.

HW3 (b) - Convergence analysis of dual variable.

$$\|\lambda^{k+1}-\lambda^*\|_2^2 = \|\lambda^k + \alpha_k(Aw_k - b) - \lambda^*\|_2^2.$$

$$= \|\lambda^k-\lambda^*\|_2^2 + 2\alpha_k\langle Aw_k - b, \lambda^k-\lambda^*\rangle + \alpha_k^2\|Aw_k - b\|_2^2$$

f: u-strong convex L-smooth $\Rightarrow$ g $\frac{1}{L}$ - strong convex and $\frac{1}{u}$ - smooth.

there is, $\|\lambda^{k+1}-\lambda^*\|_2^2 \le (1-\frac{\alpha_k}{L})\|\lambda^k-\lambda^*\|_2^2 - 2\alpha_k(g(\lambda^k)-g(\lambda^*))$
$$+\alpha_k^2\|Aw_k - b\|_2^2$$

(By using $\frac{1}{L}$-strong convexity).

Further, there is $\|\lambda^{k+1}-\lambda^*\|_2^2 \le (1-\alpha_k/L)\|\lambda^k-\lambda^*\|_2^2 - 2\alpha_k(g(\lambda^k)-g(\lambda^*))$
$$+\frac{2\alpha_k^2}{u}(g(\lambda^k)-g(\lambda^*)).$$

$$=(1-\alpha_k/L)\|\lambda^k-\lambda^*\|_2^2 - 2\alpha_k(1-\frac{\alpha_k}{u})(g(\lambda^k)-g(\lambda^*))$$

Convergence guaranteed when

~~$0<1-\alpha_k/L<1$ and~~ $1-\frac{\alpha_k}{u}\ge 0 \Rightarrow \alpha_k \le u$

and $0<1-\frac{\alpha_k}{L}<1$ since $\alpha_k\le u$

When then require $L>u$

And the convergence rate is thus $\frac{u}{L}$

HW. 3-c

$$\underset{w^1, \dots, w^N \in W}{\text{minimize}} \quad \sum_{i=1}^{N} f^i(w^i)$$

$$\text{s.t.} \quad w^1 = w^2 = \dots = w^N$$

where there are N numbers of nodes.

- Let w^i_j represents node i's estimate of nodes j's internal state.
- Let λ_{ij}, $i, j \in \{1, 2, \dots, N\}$, be the dual variables.

For instance λ_{ij} is associated with the constraint

$$w^i_i = w^j_i$$

Let

$$\lambda = \left[\lambda_{11}^T, \dots, \lambda_{1N}^T, \lambda_{21}^T, \dots, \lambda_{2N}^T, \dots, \lambda_{NN}^T\right]^T$$

The Langrangian :

$$L(w^1, w^2, \dots, w^N, \lambda) = \sum_{i=1}^{N} f^i(w^i_1, w^i_2, \dots, w^i_N)$$
$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{ij}^T (w^i_i - w^j_i)$$

The dual function

$$g(\lambda) = \inf_{w^1, \dots, w^N \in W} L(w^1, \dots, w^N, \lambda)$$

We can write the dual function in subproblems such as

$$\phi^i(\lambda) = \inf_{w^i \in W} f^i(w^i_1, \dots, w^i_N) + \sum_{j=1}^{N} \lambda_{ij}^T . w^i_i - \sum_{j=1}^{N} \lambda_{ji}^T w^i_j$$

Since $\phi^i(\lambda)$ only depends on λ and w^i, node i can compute $\phi^i(\lambda)$ locally.

The Langrangian is the sum

$$g(\lambda) = \sum_{i=1}^{N} \phi^i(\lambda)$$

Finally, the dual problem is the maximization

$$g^* = \max_{\lambda} g(\lambda) = \max_{\lambda} \sum_{i=1}^{N} \phi^i(\lambda)$$

Communication cost

Suppose that there are N nodes, each part is transferred over N-1 nodes. Summing over all parts will give the the communication cost for primal method

$$\sum_{i=1}^{N} (N-1) \dim(w_i^i) = (N-1) \dim(w)$$

Similarly, the communication cost for dual method is

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (N-1) \dim \lambda_{ij} = (N-1) \dim(\lambda)$$

Since $\dim(\lambda) > \dim(w)$, the communication cost of the dual method is higher.

But, one can decrease this amount since a node only needs the sum of dual variables, $\sum_{j=1}^{N} \lambda_{ij}$, for each iteration.