

### Homework 3 (Group 5)

(discuss)

HW3 (a)

$$\min f(w)$$

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

Define dual function  $g(\lambda) = \inf_w L(w, \lambda) := f(w) + \lambda^T(Aw - b)$

(For any given  $\lambda$ , denote  $w(\lambda) = \arg \inf_w L(w, \lambda)$ )

Show that if  $w \in W(\lambda)$ , then  $Aw - b \in \partial g(\lambda)$

$$(= \inf_w L(w, \lambda) \Rightarrow \inf_w L(w, \lambda))$$

proof: For any arbitrary  $\bar{\lambda}$ , we have that:  $\bar{\lambda} = \lambda + \Delta\lambda$

$$g(\bar{\lambda}) = \inf_w \{f(w) + \bar{\lambda}^T(Aw - b)\}$$

$$\leq f(w) + \bar{\lambda}^T(Aw - b)$$

$$= f(w) + (\bar{\lambda} - \lambda)^T(Aw - b) + \lambda^T(Aw - b)$$

$$= \{g(\lambda) = f(w) + \lambda^T(Aw - b)\}$$

$$+ \Delta\lambda^T(Aw - b) + (\Delta\lambda)^T(Aw - b)$$

Thus, we have shown that:

$$g(\bar{\lambda}) \leq g(\lambda) + (\bar{\lambda} - \lambda)^T(Aw - b)$$

which means  $Aw - b \in \partial g(\lambda)$  (since  $g(\lambda)$  is concave)

$$(x^* - p)^T \frac{\Delta}{\lambda} + \lambda^T w - w^* \geq \lambda^T w - p^T$$

$$(\lambda^T w - p^T) \frac{\Delta}{\lambda} + \lambda^T w - w^* \geq \lambda^T w - p^T$$

$$\lambda^T w - w^* \geq \lambda^T w - (\lambda^T w - p^T) \geq p^T - w^*$$

$$\lambda^T w - w^* \geq \lambda^T v - (\lambda^T v - w^*) \geq p^T - v^*$$

$$p^T - v^* \geq p^T - w^* \geq (p^T - w^*) - (\lambda^T v - \lambda^T w) = \lambda^T w - \lambda^T v$$

$$p^T - v^* \geq (p^T - w^*) - (\lambda^T v - \lambda^T w) \geq$$

$$p^T - v^* \geq (p^T - w^*) - (\lambda^T v - \lambda^T w) \geq p^T - v^*$$

HW3(b)

(2 points)

& discussion

Analyze the convergence of dual ascent for  $L$ -smooth and  $\mu$ -strongly convex  $f$ . Is the solution primal feasible?

(with proof)

dual ascent

Solution: For this problem since  $f$  is strongly convex, it holds that  $f^*$  is differentiable. Thus  $g(\lambda)$  is differentiable and  $Aw - b = \nabla g(\lambda)$ .

Hence the following algorithm

$$w_{k+1} \in \operatorname{argmin}_w L(w, \lambda_k)$$

$$\lambda_{k+1} = \lambda_k + \alpha_k(Aw_k - b)$$

turns out to be dual gradient ascent.

In order to analyze the convergence of the above algorithm, firstly the property of  $g(\lambda)$  should be investigated.

Claim: if  $f$  is strongly convex and  $x^*$  is its minimizer, then

$$f(y) \geq f(x^*) + \frac{\mu}{2} \|y - x^*\|^2 \quad \forall y \in \text{dom } f \quad \text{--- (*)}$$

Denote  $F_u(x) = f(x) - u^T x$ . Then we have the following properties:

①  $f(x)$  is  $\mu$ -strongly convex  $\Rightarrow F_u(x)$  is  $\mu$ -strongly convex

② Denote  $x_u \triangleq \nabla f^*(u)$ . Since  $\nabla f^*(u) = \operatorname{arg\sup}_w (u^T w - f(w))$ , we

know that  $x_u$  is the unique minimizer of  $F_u(x)$ .

Then by (\*) it holds that:

$$f(y) - u^T y \geq f(x_u) - u^T x_u + \frac{\mu}{2} \|y - x_u\|^2 \quad \forall y$$

Denote  $F_v(x) = f(x) - v^T x$  and  $x_v \triangleq \nabla f^*(v)$ , the same properties hold.

Thus we have:

$$\begin{cases} f(x_v) - v^T x_v \geq f(x_u) - u^T x_u + \frac{\mu}{2} \|x_v - x_u\|^2 \\ f(x_u) - v^T x_u \geq f(x_v) - v^T x_v + \frac{\mu}{2} \|x_u - x_v\|^2 \end{cases}$$

Combining the above equalities, it holds that:

$$\mu \|x_u - x_v\|^2 \leq (x_u - x_v)^T (u - v) \leq \|x_u - x_v\| \|u - v\|$$

$$\Rightarrow \|\nabla f^*(u) - \nabla f^*(v)\| \leq \frac{1}{\mu} \|u - v\|$$

Thus,  $f^*$  is  $\frac{1}{\mu}$ -smooth, which means that  $-g(\lambda)$  is also  $\frac{1}{\mu}$ -smooth.

Now consider the problem

$$\max g(\lambda)$$

where  $-g(\lambda)$  is convex and  $\frac{1}{\mu}$ -smooth.

We know that the gradient ascent method with constant step-size  $\alpha \leq \mu$  converges at rate  $O(\frac{1}{t})$ .

In addition, since the primal problem is convex with linear equality constraint, we know that strong duality holds.

Thus, the optimal solution of the dual problem also yields primal feasibility.

Hw3 (c)

$$(P2) \min \frac{1}{N} \sum_{i=1}^N f_i(w_i)$$

s.t.  $w_i = w_j$  for all  $j \in N_i$ ,  $i = 1, \dots, N$

Solution: The constraint of local consensus is equivalent to

$$a_{ij}(w_i - w_j) = 0$$

for some doubly stochastic matrix  $A = [a_{ij}]$  compatible with  $G$ .

Note that  $a_{ij} = a_{ji}$  and  $a_{ij} = 0$  if  $(i, j) \notin E$ .

Then (P2) is equivalent to

$$(P2') \min \frac{1}{N} \sum_{i=1}^N f_i(w_i)$$

s.t.  $a_{ij}(w_i - w_j) \geq 0 \quad \forall i, j = 1, \dots, N$

Then the Lagrangian for (P2') is

$$\begin{aligned} L(w, \lambda) &= \frac{1}{N} \sum_{i=1}^N f_i(w_i) + \sum_{i=1}^N \sum_{j \neq i} \lambda_i^{(i)} a_{ij}(w_i - w_j) \\ &= \sum_{i=1}^N \left[ \frac{1}{N} f_i(w_i) + \sum_{j \neq i} \lambda_j^{(i)} a_{ij} w_i - \sum_{j \neq i} \lambda_j^{(i)} a_{ji} w_i \right] \\ &= \sum_{i=1}^N \underbrace{\left[ \frac{1}{N} f_i(w_i) + a_{ii} \sum_{j \neq i} (\lambda_i^{(i)} - \lambda_j^{(i)}) w_i \right]}_{\triangleq L_i(w_i, \lambda)} \end{aligned}$$

(Note: the number of Lagrangian multipliers coincides with the number of constraints)

Thus,  $L(w, \lambda)$  is separable in  $w$  and dual decomposition can be applied.

### Dual Decomposition Algorithm

- step 1 (primal update)

$$w_{i,k+1} \in \operatorname{argmin}_{w_i} L_i(w_i, \lambda_k) \quad i=1, \dots, N$$

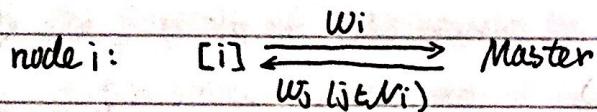
- step 2 (dual update)

$$\lambda_{i,k+1}^{(i)} = \lambda_{i,k}^{(i)} + \alpha_k a_{ij}(w_{i,k} - w_{j,k}) \quad i=1, \dots, N, j=1, \dots, N$$

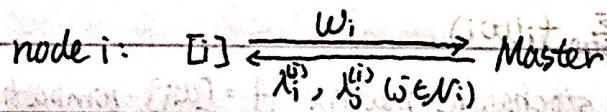
### Compare primal method and dual method

#### 1. communication cost

- primal: During each iteration, each node sends its value to the master and receives the values of its neighbors from the master.



- dual: During each iteration, each node sends its value to the master and receives the value of Lagrangian multipliers from the master.



#### Conclusion:

From the above analysis, we know that for (P2), usually the dual decomposition method requires more communication cost.

#### 2. convergence.

The convergence to a global minimum can be guaranteed for the primal method only when  $f(w)$  is convex or initial guess  $w_0$  is close to  $w^*$ .

But for the dual method, since the dual problem is always convex regardless of  $f$ , convergence can be guaranteed under more general conditions.