MLoN Homework3

Group 1

1 Problem 1

Problem

Consider

$$\begin{aligned} &\text{minimize } f(\omega) \\ &\text{s.t. } A\omega = b \end{aligned}$$

Lagrange dual function: $g(\lambda) = \inf_{\omega}\{L(\omega, \lambda) := f(\omega) + \lambda^T(A\omega - b)\}$
Lagrange dual problem: $maximize_{\lambda}\{g(\lambda) = -f^*(-A^T\lambda) - \lambda^T b\}$
Show that for convex and closed f : $A\omega - b \in \partial g(\lambda)$.

Proof

Taking the derivative of $g(\lambda)$ w.r.t. λ, we get

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

Thus, $A\omega - b \in \partial g(\lambda)$ satisfies if $\omega \in \partial f^*(-A^T\lambda)$
According to duality theorem,

$$\omega \in \partial f^*(-A^T\lambda) \Leftrightarrow -A^T\lambda \in \partial f^{**}(\omega) \Leftrightarrow \underbrace{-A^T\lambda}_{y} \in \partial f(\underbrace{\omega}_{x}) \Leftrightarrow y \in \partial f(x)$$

According to the definition of subgradients, we get

$$\begin{aligned} y \in \partial f(x) &\Leftrightarrow f(z) \geq f(x) + y^T(z - x) \\ &\Leftrightarrow < y, x > -f(x) \geq < y, z > -f(z) \\ &\Leftrightarrow < y, x > -f(x) = \sup_{z}(< y, z > -f(z)) = \inf_{z}(f(z) - < y, z >) \\ &\Leftrightarrow x \in \operatorname*{argmin}_{z}(f(z) - < y, z >) \\ &\Leftrightarrow \omega \in \operatorname*{argmin}_{z}(f(z) + \lambda^T Az) \end{aligned}$$

Thus, $A\omega - b \in \partial g(\lambda)$ satisfies when $\omega \in \operatorname*{argmin}_{z}(f(z) + \lambda^T Az)$.

2 Problem 2

Problem Analyze the convergence of dual ascent for L-smooth and μ-strong convex f. Is the solution primal feasible?

Proof Consider

$$\begin{aligned} &\text{minimize } f(\omega) \\ &\text{s.t. } A\omega = b \end{aligned}$$

Lagrange dual function: $g(\lambda) = \inf_{\omega}\{L(\omega, \lambda) := f(\omega) + \lambda^T(A\omega - b)\}$
Lagrange dual problem: $\text{maximize}_{\lambda}\{g(\lambda) = -f^*(-A^T\lambda) - \lambda^T b\}$

Dual ascent algorithm:

$$\begin{aligned} \omega_{k+1} &\in \arg\min_{w} L(w, \lambda_k) \\ \lambda_{k+1} &= \lambda_k + \alpha_k(A\omega_k - b) \end{aligned} \tag{1}$$

First take a look at the properties of $g(\lambda)$:

1. f is strongly convex, then we have f^* is differentiable, then $g(\lambda)$ is differentiable and $\nabla g(\lambda) = A\omega - b$
2. Assume x is the minimizer of f, since f is μ-strongly convex, then:

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

Assume $F_u(x) = f(x) - u^T x$, naturally we have $F(x)$ is also μ-strongly convex.

Denote $x_u \triangleq \nabla f^*(u)$, since $\nabla f^*(u) = \arg\min_{\omega}(f(\omega) - u^T\omega)$, so we can conclude that x_u is the minimizer of $F_u(x)$.

By equation (2), we have:

$$\begin{aligned} F_u(y) &\geq F_u(x_u) + \frac{\mu}{2}\|y - x_u\|^2, \forall y \\ f(y) - u^T y &\geq f(x_u) - u^T x_u + \frac{\mu}{2}\|y - x_u\|^2, \forall y \end{aligned} \tag{3}$$

The same holds for denoting $F_v(x) = f(x) - v^T x$ and $x_v \triangleq \nabla f^*(v)$

Thus we have:

$$\begin{aligned} f(x_v) - u^T x_v &\geq f(x_u) - u^T x_u + \frac{\mu}{2}\|x_v - x_u\|^2 \\ f(x_u) - v^T x_u &\geq f(x_v) - v^T x_v + \frac{\mu}{2}\|x_v - x_u\|^2 \end{aligned} \tag{4}$$

Adding those two above gives us:

$$\begin{aligned} & \mu \|x_v - x_u\|^2 \leq (x_u - x_v)^T (u - v) \\ \implies & \mu \|x_v - x_u\|^2 \leq \|x_u - x_v\| \|u - v\| \\ \implies & \|x_v - x_u\| \leq \frac{1}{\mu} \|u - v\| \\ \implies & \|\nabla f^*(u) - \nabla f^*(v)\| \leq \frac{1}{\mu} \|u - v\| \end{aligned} \tag{5}$$

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

So the Lagrange dual problem is to maximize$_\lambda\{g(\lambda)\}$, where $g(\lambda)$ is convex and $\frac{1}{\mu}$-smooth. Since we have f is μ-strongly convex and L-smooth, apply what we know about primal gradient descent, it converges at linear rate $\mathcal{O}(\log \frac{1}{\epsilon})$.

The dual problem converges, then the dual solution is feasible thus the primal solution is also feasible.

3 Problem 3

Problem

$$\text{minimize } \frac{1}{N}\sum_{i\in[N]} f_i(\omega_i)$$
$$\text{s.t. } \omega_i = \omega_j,\ \forall\ j \in \mathcal{N}_i$$

Extend the dual decomposition on connected communication graph to solve (P2) and do comparision with the primal method in terms of total communication cost and convergence rate on a random geometric communication graph.

Solution

Dual decomposition
To extend the dual decomposition to solve (P2) by modifying the algorithms in page 6-12, we need to rewrite the constraints of the problem. We know that we need to convert the equality constraint $\omega_i = \omega_j$ to the form of $\sum_{i\in[N]} A_i w_i = b$. Then we can do parallel processing in the primal update and do dual update after gathering the result from each node. There are two simple ways to represent the undirected graph connection constraints. Our first trial is constructing an $e \times N$ matrix A containing [1,0]. Each row corresponds to a certain edge and element '1' represents the node of that edge. So we have $A\omega = 0$ where ω is the weight vector. However, this is not straight forward to extend the algorithm because the worker should be placed on each edge instead of each node.

The second trial is to use the concept of the Laplacian matrix L (symmetric) which is another popular way to represent a graph. The size of L is $N \times N$. The element of L is given by:

$$L_{i,j} := \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases}$$

$\deg(v_i)$ is the degree of the vertex i. We have $L_i\omega = 0$ where L_i is ith row of L. Now we have the Lagrangian:

$$L(\omega, \lambda) = \sum_{i\in[N]} L_i(\omega_i, \lambda) = \sum_{i\in[N]} (f_i(\omega_i) + \lambda^T L_i \omega_i)$$

Lagrangian is separable in ω so we can use the primal update:

$$\omega_{i,k+1} \in argmin_{\omega_i} L_i(\omega_i, \lambda_k), i = 1, ..., N$$

In the dual update, we gather $L_i\omega_{i,k+1}$ and update λ_{k+1}.

$$\lambda_{k+1} = \lambda_k + \alpha_k \sum_{i \in [N]} L_i \omega_{i,k+1}$$

Primal method
The primal method for (P2) is described in slide 6-20.

$$\begin{aligned} \omega_{i,k+1} &= a_{ii}\omega_{i,k} - \alpha_k g_i(\omega_{i,k}) + \sum_{i \in N_i} a_{i,j}\omega_{j,k} \\ &= \sum_{i \in [N_i]} a_{i,j}\omega_{j,k} - \alpha_k g_i(\omega_{i,k}) \end{aligned}$$

Comparison on a random geometric communication graph
Definition: A random geometric graph (RGG) is an undirected geometric graph with nodes randomly sampled from the uniform distribution of the underlying space $[0,1)^d$. Two vertices $p, q \in V$ are connected if, and only if, their distance is less than a previously specified parameter $r \in (0,1)$, excluding any loops. In the 2-D graph, $d = 2$.

There are two situations: connected graph and disconnected graph. When we are dealing with the connected case, the problem is similar to the consensus problem. From slide 6-19, we know the convergence rate is $O(NlogNlog\epsilon^{-1})$. When the graph is disconnected, the convergence rate is determined by the number of the subgraph N_{sub} and also the size of each subgraph S_j. These are decided by the selected distance threshold r for the RGG. It's hard to derive the final result but we think we need extra $N_{subgraph}$ computational node for collecting the information from different subgraphs, before sending them to the master node. The communication cost will also go higher.
Intuitively, primal method will have higher communication cost because every node need to know the ω information from the neighbours.