EE61012 – Spring Semester 2024

Homework 3: Convex Optimization in Control and Signal Processing

Prof. Ashish Ranjan Hota
Department of Electrical Engineering, IIT Kharagpur

Q 3.1: ML Estimation

Let Y be a randon variable with Poission distribution having parameter μ. The probability mass function of Y is given by $\mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k!}$ for $k = 0, 1, 2, \ldots$. Let $\mu = \sum_{i=1}^{k}(\theta_i u_i) + \theta_{k+1}$ where u_i are features/contexts and $\theta \in \mathbb{R}^{k+1}$ are the unknown weights. We have access to N samples in the form of $(\widehat{u}^i, \widehat{y}^i)_{i\in\{1,2,\ldots,N\}}$. Formulate the problem of finding maximum likelihood estimate of θ as an optimization problem, and determine if the problem is convex.

Q 3.2: GD for Strongly Convex Functions

Recall that the gradient descent algorithm for an unconstrained convex optimization problem is given by

$$x_{t+1} = x_t - \eta_t \nabla f(x_t).$$

Let the cost function f be α-strongly convex and $||\nabla f(x)|| \leq G$ for all x. Then, if $\eta_t = \frac{1}{\alpha(t+1)}$, then we have

$$\frac{1}{T}\sum_{t=0}^{T-1} f(x_t) - f(x^\star) \leq \frac{G^2 \log(T)}{2T\alpha}.$$

Hint: Use potential function $\Phi_t = \frac{t\alpha}{2}||x_t - x^\star||^2$.

Let $\lambda_t := \frac{2t}{T(T+1)}$ and $\bar{x}_t := \sum_{t=1}^{T} \lambda_t x_t$. Then, under the above algorithm, we have

$$f(\bar{x}_T) - f(x^\star) \leq \frac{G^2}{\alpha(T+1)}.$$

Q 3.3: Inequality

Let $\gamma = \frac{1}{\kappa-1}$. Show that $(1+\gamma)^{-T} \leq e^{-\frac{T}{\kappa}}$.

Q 3.4: Projection is Non-Exapansive

Let $\Pi_X(y)$ denote the projection of a point y on a set X. Show that

$$||\Pi_X(y_1) - \Pi_X(y_2)||_2 \leq ||y_1 - y_2||_2.$$

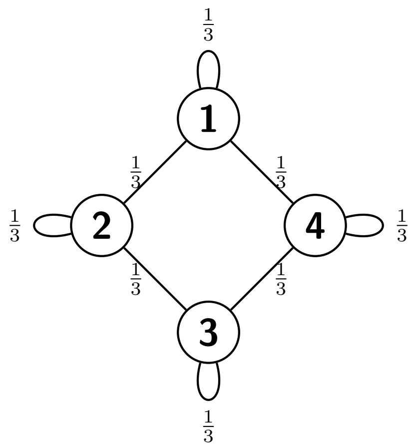

Figure 1 (for T 3.7): Network Topology for the Distributed Least Squares Problem

Q 3.5: Convergence of AGD

Prove Theorem 6.

Q 3.6: Least Squares Regression

A dataset containing 100 samples in the form $\{x^i, y^i\}_{i\in[N]}$ where each $x^i \in \mathbb{R}^3$ and $y^i \in \mathbb{R}$ is given. The output y is a polynomial of degree at most 2 of the input x. Answer the following questions.

1. Define a suitable feature map $\phi(x)$ which maps $x \in \mathbb{R}^3$ to entries of a polynomial of degree 2 in x. What is the dimension of $\phi(x)$?
2. Formulate a least squares problem to determine the coefficients of this polynomial. Clearly state the decision variable w, its dimension, and the cost function.
3. Compute the gradient of this cost function with respect to the decision variable w.
4. Find the optimal weights $w^\star$ using a suitable solver.
5. Compute the error vector $y^i - \phi(x^i)^\top w^\star$ and plot its histogram.
6. Solve the above problem using Gradient Descent, Accelerated Gradient Descent and Stochastic Gradient Descent for 1000 steps. Plot the value of the cost function computed at each iteration and the error $||w_t - w^\star||$ vs. number of interations for all three algorithms.

Hint: For a quadratic function $x^\top Ax$, its smoothness parameter is the largest eigenvalue of A and the strong convexity parameter is the smallest eigenvalue of A.

Q 3.7: Distributed Least Squares Regression

Consider a distributed version of the above least squares problem where there are four agents $\{1, 2, 3, 4\}$. The communication topology among the agents and the weights between any two neighboring agents are shown in Figure 1. Suppose the above dataset is held by the agents as follows: the first 20 data points are with agent 1, the next 25 data points are with agent 2, the next 30 data points are with agent 3 and the final 25 data points are with agent 4. Use distributed gradient descent for 100000 steps to solve the regression problem.

1. Initialize a local solution w_i^0 for each agent i randomly.
2. Determine the weights a_{ij} values for each pair of nodes.
3. Choose step size $\eta_t = \frac{1}{2\beta\sqrt{t}}$ where β is the smoothness parameter.
4. Plot $||w_i^t - \bar{w}^t||^2$ for each agent i (in log scale) and show convergence of these quantities to 0.
5. Plot $||w^\star - \bar{w}^t||^2$ and show convergence of this quantity to 0.