Advanced Optimization

Homework 3

Autumn 1400
Due date: Dey 3th

Analytical Questions

1. Show by example that in the Manifold Suboptimization, if several constraints $\mathbf{a_j^\intercal x} \leq \mathbf{b_j}$ with negative values $\mu_\mathbf{j}$ are simultaneously dropped, then the corresponding vector $\mathbf{\bar{p}_k}$ need not to be a feasible direction.

2. Let $\mathbf{H^k}$ be a positive definite matrix and define $\mathbf{\bar{x}^k}$ by

$$\mathbf{\bar{x}^k} = \arg\min_{\mathbf{x}\in\mathbf{X}} \left\{ \nabla \mathbf{f}\left(\mathbf{x^k}\right)^\intercal \left(\mathbf{x} - \mathbf{x^k}\right) + \frac{\mathbf{1}}{\mathbf{2s^k}} \left(\mathbf{x} - \mathbf{x^k}\right)^\intercal \mathbf{H^k} \left(\mathbf{x} - \mathbf{x^k}\right) \right\}.$$

Show that $\mathbf{\bar{x}^k}$ solves the problem

$$\begin{aligned} \min \quad & \frac{\mathbf{1}}{\mathbf{2}} \left\| \mathbf{x} - \left(\mathbf{x^k} - \mathbf{s^k}\left(\mathbf{H^k}\right)^{-1} \nabla \mathbf{f}\left(\mathbf{x^k}\right)\right) \right\|^{\mathbf{2}}_{\mathbf{H^k}} \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X}, \end{aligned}$$

where $\|.\|_{\mathbf{H^k}}$ is the norm defined by $\|\mathbf{z}\|^{\mathbf{2}}_{\mathbf{H^k}} = \mathbf{z^\intercal H^k z}$.

Computer Questions

1. Solve the below 3-dimensional quadratic problem, using Manifold Suboptimization method for the constraint sets a and b, and once again by using two-metric projection method for the last constraint set c. Start from the initial point $\mathbf{x^0} = [\mathbf{0\ 0\ 1}]^\intercal$.

$$\mathbf{min} \quad \mathbf{x_1^2 + 2x_2^2 + 3x_3^2}$$

   (a) $\mathbf{x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_1 + x_2 + x_3 \geq 1}$
   (b) $\mathbf{x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_1 + x_2 + x_3 \geq 1,\ x_1 \leq 0.5}$
   (c) $\mathbf{x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0}$

2. Minimize the following function,

$$\mathbf{f}\left(\mathbf{x_1}, \mathbf{x_2}\right) = \left(\mathbf{x_2} - \mathbf{0.129x_1^2} + \mathbf{1.6x_1} - \mathbf{6}\right)^{\mathbf{2}} + \mathbf{6.07}\cos\left(\mathbf{x_1}\right) + \mathbf{10},$$

with the Trust-Region method. Use $\mathbf{x^0} = [\mathbf{6\ 14}]^\intercal$ as starting point and $\mathbf{\Delta_0 = 2}$, $\mathbf{\hat{\Delta} = 5}$ as Trust-Region radius constants. Solve the sub-problems using Cauchy point and Dog-Leg methods then compare the difference in performance. Report the results thoroughly.

3. As you know from the previous homework, the auto-encoder network consists of an encoder and a decoder with the following cost function:

$$\mathbf{f}\left(\theta\right) = \sum_{\mathbf{q=1}}^{\mathbf{Q}} \left\| \mathbf{x_q} - \mathbf{D}\left(\mathbf{E}(\mathbf{x_q})\right) \right\|^{\mathbf{2}}$$

where

$$\mathbf{E}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{1} + \exp(-\mathbf{w^T x} - \mathbf{b_e})} \text{ and } \mathbf{D}(\mathbf{x}) = \mathbf{w^T E(x)} + \mathbf{b_d}$$

Implement the Levenberg-Marquardt method to minimize the cost function of the auto-encoder network for the MNIST database considering what follows:

- Use the Marquardt suggestion for adaptive value of λ.
- To reduce the amount of computation required, use only the train and test data associated with the digits of **0** and **1**.
- As implied from the equations, parameters of the problem are $\theta = \{\mathbf{w}, \mathbf{b_e}, \mathbf{b_d}\}$, with the corresponding dimensions $\theta = \{\mathbf{w : 1 \times 784, b_e : 1 \times 1, b_d : 784 \times 1}\}$.
- Consider zero as the initial value for all parameters.
- The *CCR* criteria for test data is shown below:

$$\mathbf{y_q^{'}} = \begin{cases} \mathbf{+1} & \mathbf{w^T x_q + b_e \geq 0} \\ \mathbf{-1} & \mathbf{w^T x_q + b_e < 0} \end{cases}$$

$$\mathbf{CCR} = \frac{\sum_{\mathbf{q=1}}^{\mathbf{Q}} \|\mathbf{y_q - y_q^{'}}\|}{\mathbf{2Q}}$$

- The report must contain the following elements:
  - A brief report of the modeling procedure.
  - Diagram of cost function value per iteration for the train data.
  - Diagram of λ values per iteration
  - Value of *CCR* criteria on test data.