Advanced Optimization

Homework 2

Autumn 1400
Due date: Azar 2nd

Computer Questions

1. Consider the linear equation $Ax = b$, with $A = [A_{i,j}]_{n\times n} = \left[\dfrac{1}{i+j-1}\right]$, $b = [1, \cdots, 1]$ and $x_0 = 0$ as starting point.

   - Implement the linear conjugate gradient method for $n = 5, 8, 12, 20$ with $r_k = 10^{-6}$ ($r_k = Ax_k - b$) as convergence criteria.
   - Report the number of iterations required before convergence for each dimension and justify the results.
     (hint: requested justification is in regards to the condition number of the matrix)

2. Consider the linear equation $Ax = b$, with $A \in \mathit{diag}(\mathbb{R}^{50}_{>0})$, $b = [1, \cdots, 1]_{50}$ and $x_0 = [0, \cdots, 0]_{50}$. Implement the conjugate gradient method for the following two different scenarios: one with eigenvalues of A having uniform distribution and the other with eigenvalues of A being clustered around three different values. Plot the $\log((x - x^*)^2)$ diagram for both scenarios in one figure and justify the difference.

3. The auto-encoder network consists of two parts, the encoder, and the decoder. For this network, the cost function to be minimized is $f(\theta) = \sum_{q=1}^{Q} \|x_q - D(E(x_q))\|$,
   where $E(x) = \dfrac{1}{1 + \exp(-w^T x - b_e)}$ and $D(x) = w^T E(x) + b_d$.

   Implement the Fletcher-Reeves and Polak-Ribiere nonlinear conjugate gradient methods with backtracking line search and also use the L-BFGS method to minimize the cost function of an auto-encoder network for the MNIST database considering what follows:

   - Using readily available optimization libraries is only allowed for the L-BFGS method.
   - To reduce the amount of computation required, use only the train and test data associated with the digits of 0 and 1.
   - During the training procedure, only use the train data.
   - As implied from the equations, parameters of the problem are $\theta = \{w, b_e, b_d\}$, and the corresponding dimensions are $\theta = \{w : 1 \times 784, b_e : 1 \times 1, b_d : 784 \times 1\}$. consider zero as the initial value for all parameters.
   - Try different values for α_0 and analyze its effect on convergence and CCR.
   - The CCR criteria for test data is shown below:

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

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

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

Analytical Questions

1. Consider the set of nonzero vectors $\{p_0, p_1, ..., p_l\}$ conjugate with respect to the symmetric positive definite matrix A, show that these vectors are linearly independent.

2. Assuming all the eigenvalues of matrix A, are constrained within $[a, b]$ interval, such that a, b are positive real numbers. Prove that the following inequality holds for the sequence of x_k generated from solving the equation $Ax = b$ with the conjugate gradient method.

$$f(x_2) \leq (\frac{b-a}{b+a})f(x_0)$$

3. For the Beale function, do the calculations necessary for three iterations of the BFGS algorithm on paper, using $x_0 = \begin{bmatrix} 2.1 \\ 0.4 \end{bmatrix}$ as the starting point to reach the global minimum point $x^* = \begin{bmatrix} 3 \\ 0.5 \end{bmatrix}$ with less than 0.01 tolerance.

$$f(x, y) = (1.5 - x + xy)^2 + (2.25 - x + xy^2)^2 + (2.625 - x + xy^3)^2$$

Hint: Use a calculator and do the math with at least four significant figures precision.