Homework 6 for Math 173A - Fall 2024

1. Perform the conjugate gradient method by hand on the problem

$$\Phi(x) = \frac{1}{2} x^T \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} x \quad - \quad \sum_{i=1}^{2} x_i,$$

   where $x \in \mathbb{R}^2$. Perform the algorithm either using version 0 or 1, where the conjugate directions are initialized and chosen algorithmically.

2. Here, we will prove the inequality used in class to prove fast convergence for strongly convex functions.

   Let $F(x)$ be a strongly convex function with constant c. Our goal is to show

$$F(x) - F(x^*) \leq \frac{1}{2c} \|\nabla F(x)\|^2 \qquad \text{for all } x \in \mathbb{R}^d. \tag{1}$$

   (a) Fix $x \in \mathbb{R}^d$ and define the quadratic function

$$q(y) = F(x) + \nabla F(x)^T (y - x) + \frac{c}{2} \|x - y\|^2.$$

   Find the y^* that minimizes $q(y)$.

   (b) Show that $q(y^*) = F(x) - \frac{1}{2c} \|\nabla F(x)\|^2$

   (c) Use the above to deduce (1).

   (d) Explain the proof technique in your own words to demonstrate understanding of what we did.

3. Indicate whether the following functions are strongly convex. Justify your answer.

   (a) $f(x) = x$

   (b) $f(x) = x^2$

   (c) $f(x) = \log(1 + e^x)$

4. **Coding question:** Let $A \in \mathbb{R}^{n \times n}$ be a diagonal matrix with diagonal entries

$$A_{ii} = i, \quad \text{i.e. the entries run from 1 to } n,$$

and let $b \in \mathbb{R}^n$ a vector with all 1 entries. Define the function

$$f(x) = \frac{1}{2}x^T Ax - b^T x.$$

We want to compare the convergence behavior of conjugate gradient (version 0 or 1) and gradient descent. Do the following for $n = 20$ and $n = 100$ with initialization $x^{(0)} = 0$.

(a) Find the optimal solution x^* by solving $Ax = b$ using a Matlab/Python linear equation solver (or by hand and hard code the answer).

(b) Program and run the gradient descent method for f with a fixed stepsize. Run the method for n iterations. You may experiment with the stepsize until you see something that works or use a stepsize dictated by a theorem in the class.

(c) Program and run the conjugate gradient (version 0 or 1) for f. Run the method for n iterations.

Plot the $f(x^{(t)}) - f(x^*)$ for both methods in the same figure. In a different figure, plot $\|x^{(t)} - x^*\|$ for both methods. If you encounter a number smaller than 10^{-16}, set it to be 10^{-16}. In both plots, make the logarithmic scale for the vertical axis. Comment on the plots.

You need to submit the code, the plots, and your comments for this question.