Homework 5 for Math 173A - Fall 2024

1. Consider the function $f(x_1, x_2) = (2x_1 - 1)^4 + (x_1 + x_2 - 1)^2$. Recall the minimizer is $(\frac{1}{2}, \frac{1}{2})$ for this problem. Starting at $x^{(0)} = (0, 0)$, perform two steps of Newton's method to find $x^{(2)}$. Show your work.

2. **Coding Question:** We'll consider the function $f(x_1, x_2) = 200(x_2 - x_1^2)^2 + (1 - x_1)^2$.

   (a) Write a computer program to run Newton's method on this problem.

   (b) Write a computer program to run Gradient Descent with a fixed step size $\mu = 10^{-3}$ on this problem.

   (c) Write a computer program to run Gradient Descent with backtracking line-search for this problem (you may set β and γ as you wish).

   (d) Starting at the same $x^{(0)}$, run each of the four algorithms and plot $\|x^{(t)} - x^*\|$ for each, in the same figure. In a separate figure, plot $f(x^{(t)}) - f(x^*)$ for each of them. Comment on the performance. Note that $x^* = (1, 1)$ for this problem.

3. **Coding Question:** We will consider the MNIST coding question from HW4. In this question, we run these questions for *differentiating 4's and 9's.* You can reuse the template from the previous homework for loading / formatting MNIST. Implement the following two methods and plot $F(w)$ per iteration for each. You need to submit (i) the code for the algorithms and plots, and (ii) the plots.

   - Gradient descent with backtracking line search. At each iteration t, initialize the step size $\mu = 10^{-1}$, and use $\gamma = 0.5$ and $\beta = 0.8$ to determine the correct $\mu^{(t)}$. Run your algorithm for at least 10,000 iterations and plot the loss curve $F(w^{(t)})$ as a function of t.
   - Gradient Descent with momentum.
   - Gradient Descent with Nesterov's acceleration.

   For gradient descent with momentum and gradient descent with Nesterov's acceleration, you can experiment with the parameters until you find something you like that works well.