

Please answer the following questions in complete sentences in a clearly prepared manuscript and submit the solution by the due date on Gradescope.

Remember that this is a graduate class. There may be elements of the problem statements that require you to fill in appropriate assumptions. You are also responsible for determining what evidence to include. An answer alone is rarely sufficient, but neither is an overly verbose description required. Use your judgement to focus your discussion on the most interesting pieces. The answer to “should I include ‘something’ in my solution?” will almost always be: Yes, if you think it helps support your answer.

Problem 0: Homework checklist

- Please identify anyone, whether or not they are in the class, with whom you discussed your homework. This problem is worth 1 point, but on a multiplicative scale.
- Make sure you have included your source-code and prepared your solution according to the most recent Piazza note on homework submissions.

Problem 1

(Adapted from Nocedal and Wright 3.1, typos are my own) Backtracking line search is a simple method that attempts to balance sufficient decrease with a long step length by testing a sequence of points:

$$\alpha = 1, 1/2, 1/4, \dots$$

and accepting the first point that satisfies sufficient decrease (which is the condition that $L(\alpha) \leq c_1 \alpha L'(0)$ for the line search function $L(\alpha)$ we discussed in class).

1. Implement a backtracking line search routine to select the step length and add this to our simple gradient descent code from the lectures. (Note, you are free to implement gradient descent using your own code too, adapting our code is only suggested for convenience.)
2. Prepare a plot of the step lengths (values of α_k that were accepted) when we run gradient descent with this line search on the Rosenbrock function starting from the point $(1.2, 1.2)$ and also the point $(-1.2, 1)$.
3. Implement Newton’s method using line search as well. (Don’t worry about singularity in the Hessian or any of the real issues that come up with Newton – just use the search direction $\mathbf{H}(\mathbf{x}_k)\mathbf{p} = -\mathbf{g}_k$.)
4. Prepare a plot of the step lengths as in part 2.
5. Discuss any notable differences or similarities.
6. Prompt an LLM to implement a line search method that satisfies the Wolfe or strong Wolfe conditions (your choice). Test your linesearch method in the previous two scenarios. Does it work? How many times does this need to query your function f and gradient g ?

Problem 2

(Exercise 4.2 in Kochenderfer and Wheeler) The first Wolfe condition requires $f(\mathbf{x} + \alpha\mathbf{p}) \leq f(\mathbf{x}) + c_1\alpha\mathbf{p}^T\mathbf{g}$. What is the maximum step length α that satisfies this condition given that $f(x) = 5 + x_1^2 + x_2^2$, $\mathbf{x} = [-1, 1]^T$, $\mathbf{p} = [1, 0]^T$, $c_1 = 10^{-4}$.

Problem 3

(Modified from Griva, Sofer, and Nash 11.5.8.) Consider the function

$$f(\mathbf{x}) = -\mathbf{q}^T\mathbf{x} + \sum_j y_j \log(\mathbf{C}^T\mathbf{x})_j$$

where \mathbf{C} is a very large matrix and computing $\mathbf{C}^T\mathbf{x}$ is expensive. Show how we can use the structure of the function to reduce the number of matrix vector products $\mathbf{C}^T\mathbf{x}$ that would be required for checking the Wolfe conditions when searching in a direction \mathbf{p} . (Hint, compute $\mathbf{w} = \mathbf{C}^T\mathbf{p}$ once, and see how to re-use it.)

Problem 4

Read section 3.5 in Nocedal and Wright on step-length selection. Then solve problem 3.13 (typos are mine).

1. In the notation we have from class where $L(\alpha)$ is the function projected into the current search direction, then show that the quadratic function that interpolates $L(0)$, $L'(0)$, and $L(\alpha_0)$ is given by (3.57), i.e.

$$\phi_q(\alpha)_q = \left(\frac{L(\alpha_0) - L(0) - \alpha_0 L'(0)}{\alpha_0^2} \right) \alpha^2 + L'(0)\alpha + \phi(0).$$

2. Let α_1 be the minimizer of this quadratic. Suppose that the sufficient decrease condition is not satisfied at α_0 . Then show that $\alpha_1 < \frac{\alpha_0}{2(1-c_1)}$.

Optional Problem 5

Ask an LLM to adapt our proof from class to use the Wolfe or Strong Wolfe conditions instead of backtracking line search. Closely analyze the resulting proof. Is it correct? Ask the LLM to develop visualizations that help explain the proof like an 3blue1brown video. Does it succeed?