ECE 509 (Spring'25): Homework #4

70 points

Problem 1 (10 points): Let $f \in \mathcal{C}^1_L(\mathbb{R}^n)$ be a continuously differentiable function with L-Lipschitz continuous gradients ∇f, and let dom f be convex. Consider the gradient descent method given by:

$$x^{(k+1)} = x^{(k)} - t^{(k)} \nabla f(x^{(k)}),$$

where $t^{(k)} > 0$ is the step size. Let $\epsilon > 0$ be a fixed number and suppose that for all k:

$$\epsilon \leq t^{(k)} \leq \frac{2-\epsilon}{L}.$$

Prove that gradient descent converges to a stationary point, meaning that

$$\nabla f(x^{(k)}) \to 0 \quad \text{as} \quad k \to \infty,$$

for any initialization $x^{(0)}$.

Problem 2 (10 points): Let $f(x) = \log(1 + e^{-a^\top x})$, where $a \in \mathbb{R}^n$. Compute the Hessian $\nabla^2 f(x)$ and show that it is positive semi-definite for all values of x.

Problem 3 (10 points): Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ be invertible. Define g as $g(x) = f(Ax + b)$ for all x and let $u_0 \in \mathbb{R}^n$ be arbitrary but fixed. A step of Newton's method applied to f at u_0 results in

$$u_1 = u_0 - (\nabla^2 f(u_0))^{-1} \nabla f(u_0).$$

Show that a step of the Newton's method applied to g at $x_0 = A^{-1}(u_0 - b)$ results in $x_1 = A^{-1}(u_1 - b)$.

This will imply that $g(x_1) = f(u_1)$, that is, the criterion values match after a Newton step. This will continue to be true at all iterations, and thus we say that Newton's method is affine invariant.

Problem 4 (40 points): Refer to your previous implementation of gradient descent for fixed step size. We will now update it to also support variable step size selection using backtracking line search. Modify your function to include backtracking by passing additional arguments α and β in the function call, which correspond to the sufficient reduction parameter and the backtracking parameter, respectively.

Modify your function to implement backtracking line search with the following logic:

- **Inputs:**
  - A function computing the gradient of the objective.
  - An initialization point $x^{(0)}$.
  - A flag specifying whether to use a fixed step size or a variable step size.
  - A step size value (for fixed step size).
  - A maximum number of iterations.

- A tolerance ϵ for the stopping criterion.
- **New inputs:**
  * α (sufficient reduction parameter in Armijo condition).
  * β (backtracking parameter).

- **Stopping Criterion:** The method should stop when

$$\|\nabla f(x^{(k)})\|_2 \leq \epsilon.$$

- **Step Size Selection:**
  - If using fixed step size, use the provided step size.
  - If using variable step size, use backtracking line search:
    1. Start with an initial step size $t = 1$.
    2. While

$$f(x^{(k)} - t\nabla f(x^{(k)})) > f(x^{(k)}) - \alpha t \|\nabla f(x^{(k)})\|^2,$$

       reduce t by multiplying it by β.
    3. Once the condition is satisfied, accept the step size.

- **Output:** The function should return the entire sequence of iterates $\{x^{(k)}\}$.

Using your updated gradient descent function, test it on the quadratic function:

$$f(x) = \frac{1}{2} x^T Q x.$$

Compare fixed step size and variable step size using the following test cases:

1. When

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

   use fixed step sizes $\alpha = 0.1$ and $\alpha = 0.5$, and compare with variable step size using backtracking with $\alpha = 10^{-4}, \beta = 0.5$.

2. When

$$Q = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix},$$

   use fixed step sizes $\alpha = 0.01$ and $\alpha = 0.05$, and compare with variable step size using backtracking with $\alpha = 10^{-4}, \beta = 0.5$.

For each case and step size strategy, generate the following carefully labeled plots:

- (a) **Contour plot of $f(x)$ with iterates:** Overlay the iterates $\{x^{(k)}\}$ on the contour lines of the quadratic function.

- (b) **Function value vs. iterations (semi-log scale):** Plot $f(x^{(k)})$ vs. iteration number k on a semi-log scale to compare convergence rates for fixed and variable step sizes. Both step size strategies should be plotted on the same graph for direct comparison.

- (c) **Gradient norm vs. iterations:** Plot $\|\nabla f(x^{(k)})\|_2$ as a function of iteration number k, again showing both step size strategies on the same plot.

Compare the convergence behavior of fixed step size and variable step size. Which one converges faster?

Submit your well-commented code, all output results including numerical values and plots, and a detailed answer to the comparison question.