## ECE 509 (Spring'25): Homework #3

## 80 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  $\nabla f$ . Consider the descent method described by:

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)},$$

where  $t^{(k)} > 0$  is the step size, and  $\Delta x^{(k)}$  is the search direction. Using the quadratic upper bound property of  $\mathcal{C}_L^1(\mathbb{R}^n)$  functions:

- 1. Show that any direction  $\Delta x^{(k)}$  forming an acute angle with  $\nabla f(x^{(k)})$  is a descent direction for an appropriate step size  $t^{(k)}$ .
- 2. Determine the range of step sizes  $t^{(k)}$  that ensures a decrease in the function value, provided the current point is not optimal.

**Problem 2 (10 points):** Let  $f(x) = ||x||_2^4$  be defined on the unit ball  $\{x \in \mathbb{R}^n : ||x||_2^2 \le 1\}$ . Prove that  $\nabla f(x)$  is Lipschitz continuous on this domain and derive the Lipschitz constant L.

**Note:** If the domain of f were unbounded,  $\nabla f(x)$  would not be Lipschitz continuous.

Hint: In your algebraic manipulations, you may need to use the reverse triangle inequality, which states:

$$||x - y|| \ge |||x|| - ||y|||.$$

**Problem 3 (60 points):** In this problem, you will implement gradient descent using a programming language of your choice, test it on two different quadratic functions, and analyze its behavior under different step sizes. Your implementation should be **well-commented**, and you must submit both your **code and the output results**, including all plots and numerical results. Implement a function for gradient descent with the following requirements:

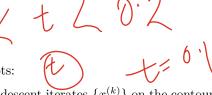
## • Inputs:

- 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  $\epsilon$  for the stopping criterion.
- Stopping Criterion: The method should stop when  $\|\nabla f(x^{(k)})\|_2 \leq \epsilon$ .
- Other Requirements:
  - Implement only fixed step size in this assignment.
  - If the input requests variable step size, the function should print: "Variable step size is currently not supported." (A future assignment will cover variable step size through line search methods.)
- Output: The function should return the entire sequence of iterates  $\{x^{(k)}\}$ .

Using your gradient descent implementation, test it on two different quadratic functions that are defined as  $f(x) = \frac{1}{2}x^TQx$ :

- 1. When  $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ :
  - Use fixed step sizes  $\alpha = 0.1$  and  $\alpha = 0.5$ .
- 2. When  $Q = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$ :
  - Use fixed step sizes  $\alpha = 0.01$  and  $\alpha = 0.05$ .

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



- (a) Contour plot of f(x) with iterates: Overlay the gradient descent iterates  $\{x^{(k)}\}$  on the contour lines of the quadratic function.
- (b) Function value vs. iterations: Plot  $f(x^{(k)})$  as a function of iteration number k.
- (c) Gradient norm vs. iterations: Plot  $\|\nabla f(x^{(k)})\|_2$  as a function of iteration number k.

Finally, answer the following questions based on your results:

- 1. How does the choice of step size affect convergence behavior?
- 2. How does changing the matrix Q affect convergence?

