Exercise 2 (15 marks) *Recall that we implemented the gradient descent algorithm to solve* $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$*. The key components in the gradient descent iterations include the descent direction* $\mathbf{p}_k$*, which is set to* $-\nabla f(\mathbf{x}_k)$*, and the step length* η_k*, determined by solving an optimization problem (or sometimes kept constant across all iterations). Finding a closed-form expression as a solution to the optimization problem for a suitable step length might not always be possible. To address general situations, we will attempt to devise a different procedure in this particular exercise. To determine the step length, we will use the following property: Suppose a non-zero* $\mathbf{p} \in \mathbb{R}^n$ *is a descent direction at point* $\mathbf{x}$*, and let* $\gamma \in (0,1)$*. Then there exists* $\varepsilon > 0$ *such that*

$$f(\mathbf{x} + \alpha\mathbf{p}) \leq f(\mathbf{x}) + \gamma\alpha\nabla f(\mathbf{x})^\top\mathbf{p}, \quad \forall\alpha \in (0, \varepsilon].$$

This condition is known as a sufficient decrease condition.

Utilizing the concept of sufficient decrease, the step length η_k *can be determined using a backtracking procedure illustrated below to find an appropriate value of* ε*.*

Algorithm 2 Backtracking (Inexact) Line Search

Require: $x_k, p_k, \alpha_0, \rho \in (0,1), \gamma \in (0,1)$

1: Initialize $\alpha = \alpha_0, p_k = -\nabla f(x_k)$
2: **while** $f(x_k + \alpha p_k) > f(x_k) + \gamma\alpha\nabla f(x_k)^\top p_k$ **do**
3: $\quad\alpha = \rho\alpha$
4: **Output:** α

This is known as approximate (inexact) line search method to find the step length at each iteration.

1. *Consider the function* $g(\mathbf{x})$ *from Exercise-1 for this part and with the starting point* $\mathbf{x}_0 = (100, 100)$ *and* $\tau = 10^{-10}$*, we will investigate the behavior of the backtracking line search algorithm for different choices of* α_0*. Set* $\gamma = \rho = 0.5$ *and try* $\alpha_0 \in \{1, 0.9, 0.75, 0.6, 0.5, 0.4, 0.25, 0.1, 0.01\}$*. For each* α_0*, record the final minimizer, final objective function value, and the number of iterations taken by the gradient descent algorithm with backtracking line search to terminate. Generate a plot where the number of iterations is plotted against* α_0 *values. Provide observations on the results, and comment on the minimizers and objective function values obtained for different choices of* α_0*. Check and comment if, for any* α_0 *value, gradient descent with backtracking line search takes a lesser number of iterations compared to the gradient descent procedure with exact line search. Plot the level sets of the function* $g(\mathbf{x})$ *and also plot the trajectory of the optimization on the same plot for both inexact line search method and the fixed step length method of gradient descent algorithm and report your observations.*
2. *Redo (1) using the function* $f(\mathbf{x})$ *from Exercise-1 and also keep in mind the answer of the part (2) from Exercise-1.*
3. *What do you conclude from (1) and (2) regarding these two line search approaches ?*