International Institute of Information Technology, Hyderabad
CS1.404 (Spring 2022): Optimization Methods
Assignment 3
Maximum Marks : 35
Submission Deadline : 5^{th} May, 2022

General Instructions to the students:

1. Attempting all questions is mandatory.
2. Marks for each of the question are mentioned at the question itself.
3. You are expected to solve all the questions using python programming language.
4. Use of any in-built libraries to solve the problem directly is not allowed.
5. Submission Format: Submit a zip file with the name rollno_Assignment2 containing the files 1.py, 2.py, 3.py, 4.py corresponding to codes for Question 1, 2, 3, 4 respectively.
6. Plagiarism is a strict No. We will pass all codes through the plagiarism checking tool to verify if the code is copied from somewhere. In that case, you get **F** grade in the course.
7. If any two students codes are found exactly same (if they copy from each other), both will get **F** grade.

Problem 1. [10 Marks] Consider minimization of the following function.

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$

Implement steepest gradient descent to minimize the function. Use the stopping condition as $\|\nabla f(x_1^k, x_2^k)\| \leq \epsilon$ with $\epsilon = 10^{-6}$. Use the following line search methods.

1. Use exact line search method. [5 Marks]
2. Backtracking line search. For backtracking line search, see the Figure 1. [5 Marks] Use

Backtracking-Armijo Line Search:

1. Given $\alpha_{\text{init}} > 0$ (e.g., $\alpha_{\text{init}} = 1$), let $\alpha^{(0)} = \alpha_{\text{init}}$ and $l = 0$.

2. Until $f(x^k + \alpha^{(l)} p^k) \leq f(x^k) + \alpha^{(l)} \beta \cdot [g^k]^\top p^k$,

   i) set $\alpha^{(l+1)} = \tau \alpha^{(l)}$, where $\tau \in (0, 1)$ is fixed (e.g., $\tau = \frac{1}{2}$),

   ii) increment l by 1.

3. Set $\alpha^k = \alpha^{(l)}$.

Figure 1: Backtracking Line Search (to be used in Problem 1). $\mathbf{p}^k$ is same as $\mathbf{d}^k$.

$\tau = 0.7$, $\beta = 0.1$.

For each case, report the following results.

1. Plot $\mathbf{x}^k$ vs $f(\mathbf{x}^k)$. In how many iterations, the algorithm converges.
2. Plot the contours of f. Plot the iterates $\mathbf{x}^k$ generated by the gradient method (shown as small circles). After every update, using arrow show the movement between successive iterates in the contour plots.

Problem 2. [5 Marks] Consider minimization of the following function.

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$

Implement Newton method to minimize the function. Use the stopping condition as $\|\nabla f(x_1^k, x_2^k)\| \leq \epsilon$ with $\epsilon = 10^{-6}$. Use backtracking line search. For backtracking line search, see the Figure 1. Use $\tau = 0.7$, $\beta = 0.1$.

Report the following results.

1. Plot $\mathbf{x}^k$ vs $f(\mathbf{x}^k)$. In how many iterations, the algorithm converges.

2. Plot the contours of f. Plot the iterates $\mathbf{x}^k$ generated by the gradient method (shown as small circles). After every update, using arrow show the movement between successive iterates in the contour plots. Also, in the same figure, plot $\{\mathbf{x} | (\mathbf{x} - \mathbf{x}^k)^T \nabla^2 f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) \leq 1\}$.

Problem 3. [10 Marks] Consider steepest descent algorithm applied to the function $f(x_1, x_2) = 5x_1^2 + 5x_2^2 - x_1 x_2 - 11x_1 + 11x_2 + 11$. Find the eigen values of the Hessian matrix. Let $\lambda_{\max}$ be the largest eigen value of the two.

1. Use fixed step size $0 < \alpha < \frac{2}{\lambda_{\max}}$. Try five different starting points. For each starting point:
   - Show contour plot of the function. After every update, using arrow show the movement in the contour plots.
   - How many iterations does it take to converge?

   [5 Marks]

2. Use fixed step size $\alpha > \frac{2}{\lambda_{\max}}$.
   - Show contour plot of the function. After every update, using arrow show the movement in the contour plots.
   - How many iterations does it take to converge?

   [5 Marks]

What do you conclude with this exercise?

Problem 4. [10 Marks] Recall that $\mathbf{x}^*$ is a local maxima of f if the function value does not increase (decrease for local minima) when you move very small distance from $\mathbf{x}^*$ in any direction. We call $\mathbf{x}^*$ a saddle point of f if the function value increases when we move from $\mathbf{x}^*$ in some direction and decreases when we move from $\mathbf{x}^*$ in another direction. Let $\mathbf{d}_\theta = [\cos(\theta), \sin(\theta)]$ donote the direction vector with angle $\theta \in [0, 2\pi]$.

1. Consider the function $f(x) = 10x_1^2 + 10x_1 x_2 + x^2 + 4x_1 - 10x_2 + 2$ over $[-3, 3] \times [-3, 3]$. Plot the function and see whether it has a local maxima / minima / saddle point at $\mathbf{x}^* = (1.8, -4)$. For clarity, plot $[f(\mathbf{x}^* + \alpha \mathbf{d}_\theta) - f(\mathbf{x}^*)]$ vs θ for $\alpha = 0.01$ and see whether it is always non-negative (if local minima), always non-positive (if local maxima) or has both positive and negative (if saddle point). Also, calculate $\nabla f(\mathbf{x}^*)$ and eigen values of $\nabla^2 f(\mathbf{x}^*)$. [5 marks]

2. Repeat part (i) with $f(\mathbf{x}) = 16x_1^2 + 8x_1 x_2 + 10x_2^2 + 12x_1 - 6x_2 + 2$, $\mathbf{x}^* = (-0.5, 0.5)$.[5 marks]

Ans: Attach the plots and fill up the following table:

	$\nabla f(\mathbf{x}^*)$	eigen values of $\nabla^2 f(\mathbf{x}^*)$	at $\mathbf{x}^*$ local maxima/minima/saddle point?
Q4.(1)			
Q4.(2)			