Monsoon Semester (Aug-Nov)
Topics in Applied Optimization (TAO)

Tutorial II

Instructor: Dr. Pawan Kumar

20 marks.

INSTRUCTIONS:

The codes can be written in either C, C++, MATLAB, Octave, FORTRAN, or Python. Write useful comments, and do proper indentation. Solution to the assignment problems will be discussed in Tutorials

1. **(Multivariable Functions)**

   1. (**Numerical Gradient**) Given a multivariable single valued function

$$f(x_1, x_2, \dots, x_n) : \mathbb{R}^n \to \mathbb{R},$$

the numerical derivative with respect to the ith variable is defined as follows:

$$\frac{\partial f}{\partial x_i} = \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}.$$

Write a code in your favourite programming language to compute numerical derivatives. For the following functions, check whether the numerical derivative and the analytical derivates match by showing plots in the indicated domain. You have to compare the plots of the partial derivatives.

(a) $f(x, y) = x^2 + y^2, \quad x, y \in [-1, 1] \times [-1, 1]$. Consider 1000 points along X and Y axes.

(b) $f(x, y, z) = \sin(x) + xz, \quad x, y, z \in [-\pi, \pi]$. Consider 1000 points along X, Y, Z axes.

To use second order methods (such as Newton's method), we need second order derivatives. The second order derivatives are given as follows:

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{f(x_1, x_2, \dots, x_i + 2h, x_{i+1}, \dots, x_n) - 2f(x_1, x_2, \dots, x_i + h, \dots, x_n) + f(x)}{h^2}$$

$$\frac{\partial^2 f}{\partial x_i x_j} = \frac{f(\dots, x_i + h, \dots, x_j + h, \dots) - f(\dots, x_i + h, \dots,) - f(\dots, x_j + h, \dots,) + f(x)}{h^2},$$

where in the last terms of both expressions, we have $f(x) = f(x_1, x_2, \dots, x_n)$. The Hessian is defined as follows:

$$H = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,n} \\ h_{2,1} & h_{2,2} & \dots & h_{2,n} \\ \vdots & \dots & \dots & \vdots \\ h_{n,1} & \dots & \dots & h_{n,n} \end{bmatrix},$$

where

$$h_{i,j} = \frac{\partial^2 f}{\partial x_i x_j}.$$

Write a program that computes the hessian at the indicated point for the following functions. Compute Hessian numerically as well as by hand and check whether they are same.

(a) $f(x, y) = x^2 + y^3$ at the point $(1, 2)$. 2

(b) $f(x, y) = \dfrac{1}{1 + \exp(3x + 2y)}$ at point $(1, 1)$. 3

(c) Consider the following cost function from logistic regression. 5

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}\log(h_\theta(x^{(i)})) - (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))],$$

where

$$h_\theta(x) = g(\theta^T x), \quad g(z) = \frac{1}{1 + \exp(-z)}, \quad z = \theta^T x,$$

also, here $x^{(i)}$ is the ith row of $m \times n$ matrix, and y_i is the ith entry of label vector y, which is of length m. Here X is called the data, and y is called the label. Here θ is the unknown vector, and it is of length n. Compute the gradient of J. That is compute:

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial \theta_1} \\ \frac{\partial J}{\partial \theta_2} \\ \vdots \\ \frac{\partial J}{\partial \theta_n}. \end{bmatrix}$$

[Note: In the future assignments, you will write a gradient method for logistic regression.]

2. 1. (**Level Curves:**) 5

(a) Level curves are a nice way to understand the properties for a 3D function (surface plots) by studying curves on 2D plane. For the level curve shown in Figure 1, Answer the following:

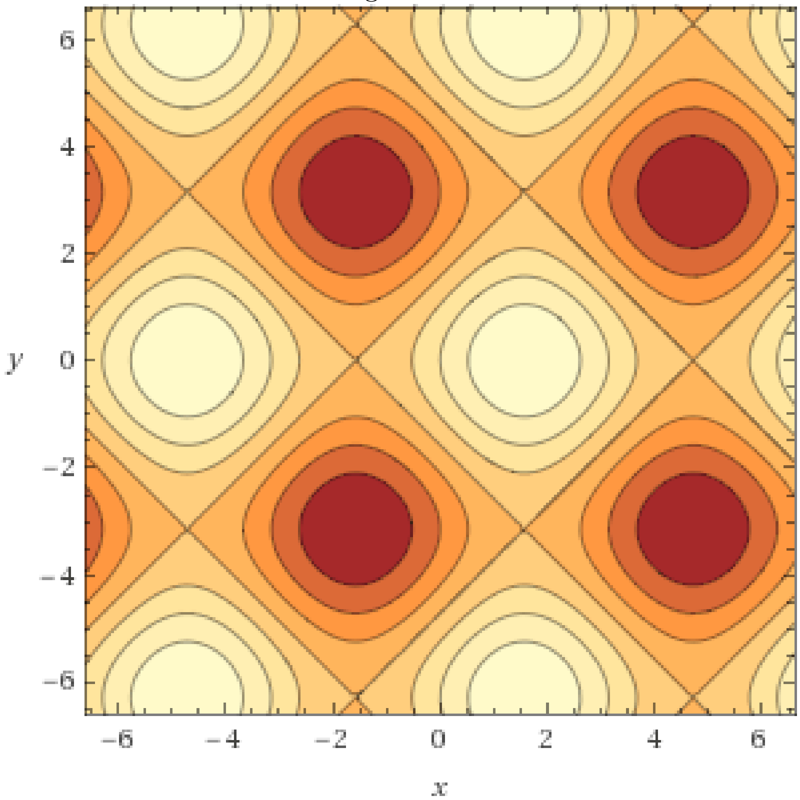
i. Guess the approximate coordinates (x, y) which are the points of either maxima or minima.

ii. Is this function periodic? If yes, then guess the approximate period.

iii. Can this be a convex function? Note that convex functions are functions with only one minima or maxima.

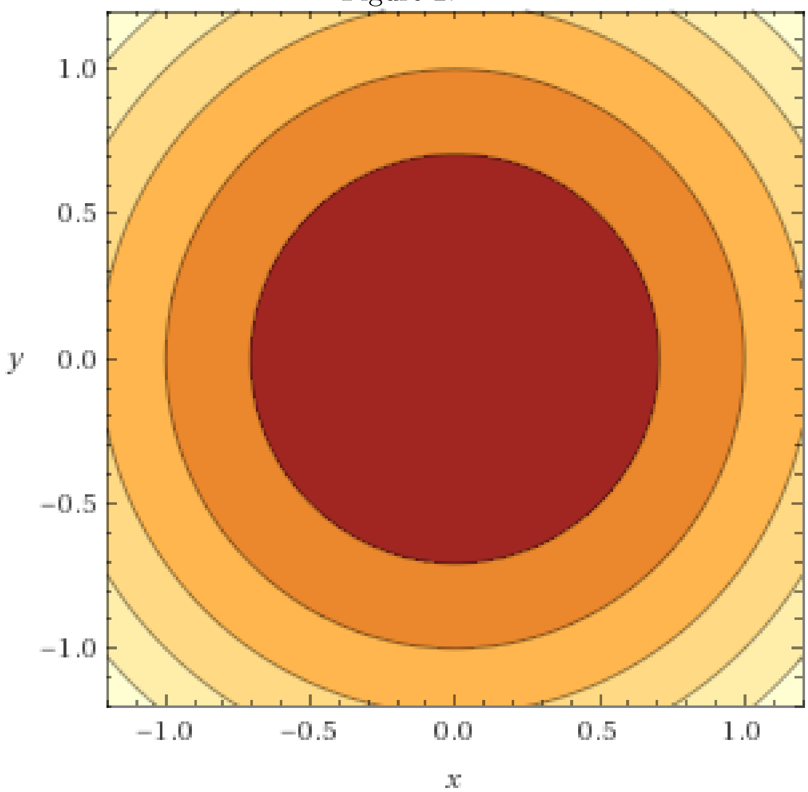
iv. If you want to run the gradient method to find one (local) minima, guess a good starting point such that gradient method will converge quickly to that local minima.

Figure 1:



Question 2 continues on the next page. . .

Figure 2:



v. Draw a sketch of a same level curves on paper, and indicate the direction of gradient vectors. Recall that gradient vector direction is orthogonal to the tangent drawn on level curves.

(b) The level curves of the function

5

$$f(x, y) = x^2 + y^2$$

in the domain (box) $[-1, 1] \times [-1, 1]$ is shown in Figure 2. Now answer the following:

i. How many minima or maxima are there? How many minima and maxima are there for the function

$$f(x, y) = (x - a)^2 + (y - b)^2?$$

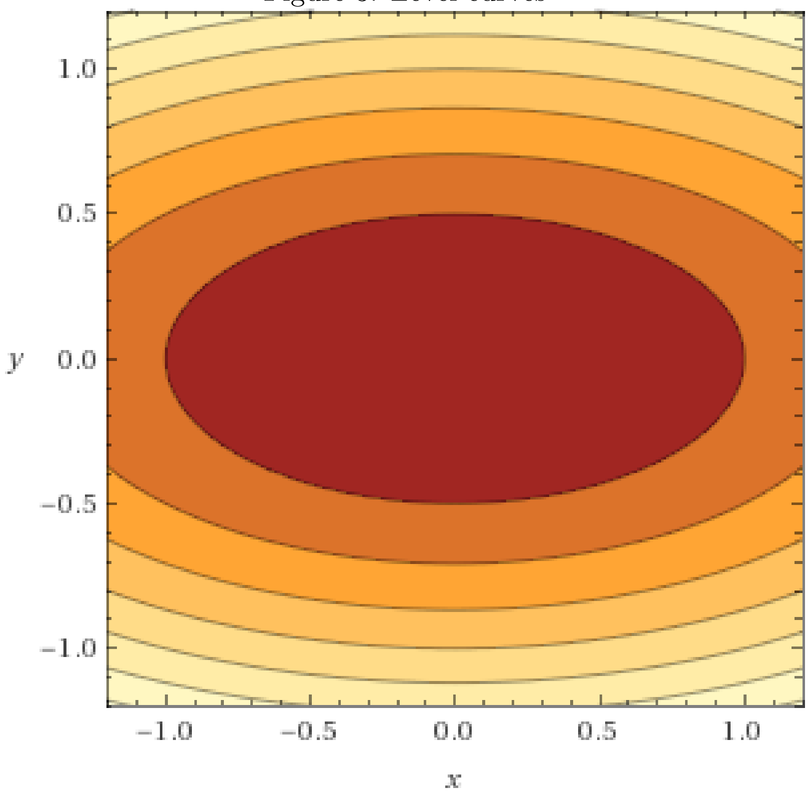
ii. How many minima or maxima will be there for the following function

$$f(x, y) = 10.99 * ((x - a)^2 + (y - b)^2)?$$

iii. If we want to run the gradient method, which is a good starting guess for a starting vector, given that we don't know minima or maxima in advance? Illustrate your answer by drawing a sketch of these level curves, and show the way you proceed to minima.

iv. The function $f(x, y) = x^2 + y^2$ can be written as $f(x, y) = [x, y]A[x, y]^T$, where A is an identity matrix. Let us now consider the following level curve in Figure 3.

Figure 3: Level curves



These are the level curves of $f(x,y) = [x,y]A[x,y]^T$. Guess A. For this function, suggest a best (fastest to minima) and worst (slowest to minima) starting guess for the gradient method by sketching these curves on paper, and showing the directions of the gradient.