ECE-6500

ASSIGNMENT - 1

- UTKARSH PRATAP SINGH JADON
500711257

Q1] $f(x) = x_1^2 + e^{x_2} + e^{-x_2} + 3x_3^4$

a) set $X = \{ x \in \mathbb{R}^3 : x_1^2 + 2x_2^2 + 3x_3^2 \leq 1 \}$

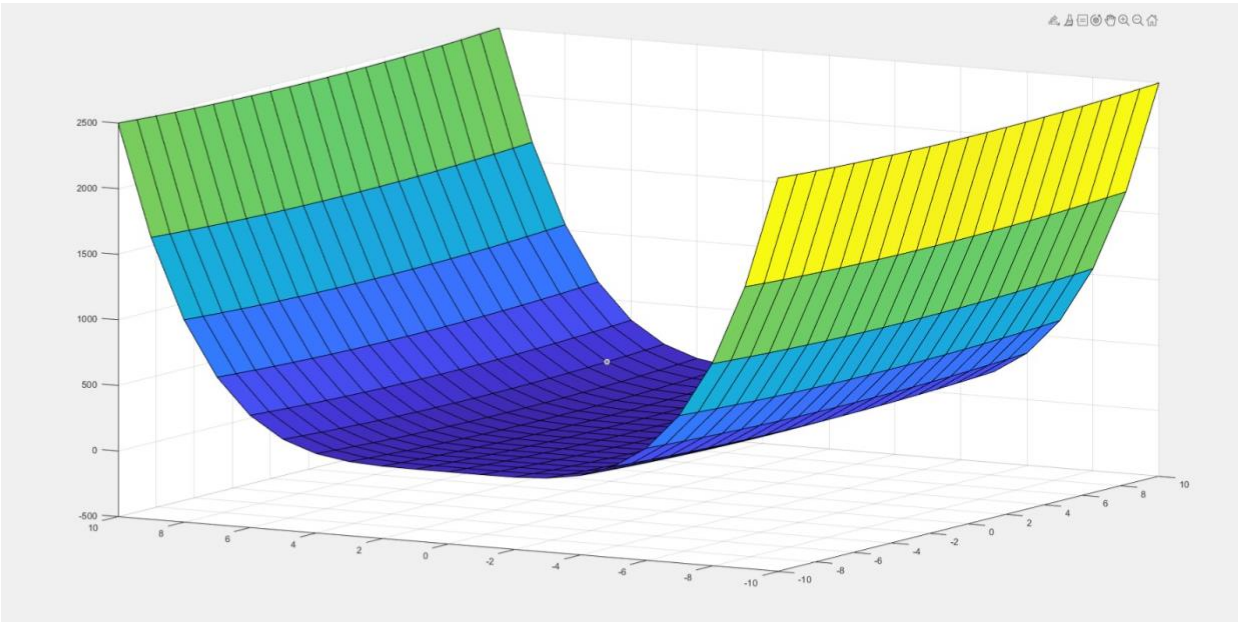
Here X is bounded and closed, hence it is compact.
$f(x)$ is continuous.
$\therefore$ Min and max both exist in X

b) $x \in \mathbb{R}^3$

Here, $\mathbb{R}^3$ is closed but not bounded
Also, $\lim_{\|x\| \to \infty} f(x) = \infty$ $\therefore$ $f(x)$ is coercive

$\therefore$ Minimum of $f(x)$ exists on $\mathbb{R}^3$
But since $f \to \infty$ as $\|x\| \to \infty$, max doesn't exist.

Q2) a.



Local minima achieved at (0,-1) where the value of function is -0.25

Q2) (c) $f(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$

$$\nabla f(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2^3 - x_2 \end{bmatrix}$$

According to gradient descent algorithm,

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

Let $\gamma = 0.1$

$\cancel{x^{k+1}_2}$ $\quad x^{(0)} = (1, 0)^T$

$$x^{(1)} = x^{(0)} - \gamma \nabla f(x^{(0)})$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 1 \\ 0^3 - 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0 \end{bmatrix}$$

Now, $x^{(2)} = x^{(1)} - \gamma \nabla f(x^{(1)})$

$$= \begin{bmatrix} 0.9 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 0.9 \\ 0 \end{bmatrix}$$

$$x^{(2)} = \begin{bmatrix} 0.81 \\ 0 \end{bmatrix}$$

$\vdots$

Thus it can be seen that $\{x^{(t)}\}_t$ will converge to zero but not to one of the identified minimas.

[Local minima : $[0, -1]$]

Q3) $f(x) = \|x\|^{3/2}$ $\quad f : \mathbb{R}^2 \rightarrow \mathbb{R}$

a) Show $f(x)$ does not have Lipschitz ∇f for any $L < \infty$

$$\cancel{\nabla f(x) =}\ f(x) = \|x\|^{3/2} = \left(\left(\sum_{k=1}^{2} x_k^2\right)^{1/2}\right)^{3/2}$$

$$\cancel{\nabla f(x) = \frac{3}{2}\|x\|^{1/2}}$$

$$= \left(\sum_{k=1}^{2} x_k^2\right)^{3/4}$$

$$\nabla_j f(x) = \frac{\partial}{\partial x_j}\left(\sum x_k^2\right)^{3/4} = \frac{3}{4}\left(\sum x_k^2\right)^{-1/4}(2x_j)$$

$$= \frac{3}{2} x_j \left(\sum x_k^2\right)^{-1/4}$$

$$\nabla f_j(x) = \frac{3}{2}\frac{x_j}{\sqrt{\|x\|}}$$

Now, testing Lipschitz condition at $y = -x$.
We know that, $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$

Substituting $y = -x$

$$\frac{3x}{2\|x\|^{1/2}} + \frac{3x}{2\|x\|^{1/2}} \leq L(2x)$$

$$L \geq \frac{3}{2\|x\|^{1/2}} \Rightarrow \|x\|^{1/2} \geq \frac{3}{2L}$$

This does not hold true for $\|x\|^{1/2} < 3/2L$

$\therefore$ Given function $f(x)$ does not have Lipschitz ∇f for any choice of $L < \infty$.

b)

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

$$\therefore \begin{bmatrix} x^{k+1} \\ y^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ y^k \end{bmatrix} - \gamma \begin{bmatrix} \dfrac{3x_k}{2(x_k^2+y_k^2)^{1/4}} \\ \dfrac{3y_k}{2(x_k^2+y_k^2)^{1/4}} \end{bmatrix}$$

From above eqⁿ, we can see that denominator converges to zero later than numerator.

$\therefore$ function will converge around 0 but not exactly 0, in finite number of iterations.

Q4) $h_\theta(x) = \theta_0 + \sum_{j=1}^{n} \theta_j x_j$

Set: $\{(x^{(m)}, y^{(m)})\}_{m=1,\dots,M}$

As per gradient descent algorithm:

$$\theta^{m+1} = \theta^{m} - \alpha \frac{\partial C}{\partial \theta}$$

where C is the Cost function, given by:

$$C = \frac{1}{2M} \sum_{m=1}^{M} \| h_\theta(x^{m}) - y^{(m)} \|^2$$

$$\frac{\partial C}{\partial \theta} = \frac{\partial}{\partial \theta}\left\{ \frac{1}{2M} \sum_{m=1}^{M} \| h_\theta(x^{m}) - y^{m} \|^2 \right\}$$

$$\frac{\partial C}{\partial \theta} = \frac{1}{M} \sum_{m=1}^{M} \left(h_\theta(x^{m}) - y^{m} \right) \cdot \frac{\partial}{\partial \theta} h_\theta(x^{m})$$

Here, $\frac{\partial}{\partial \theta} h_\theta(x^{m}) = \frac{\partial}{\partial \theta}\left(\theta_0 + \sum_{j=1}^{n} \theta_j x_j \right)$

$\therefore \frac{\partial}{\partial \theta_0} h_\theta(x^{m}) = 1 \qquad \frac{\partial}{\partial \theta_1} h_\theta(x^{m}) = x_1^{m} \qquad \frac{\partial}{\partial \theta_2} h_\theta(x^{m}) = x_2^{m}$

$\frac{\partial}{\partial \theta_n} h_\theta(x^{m}) = x_n^{m}$

$$\therefore \frac{\partial C}{\partial \theta} = \begin{bmatrix} \frac{1}{M} \sum_{m=1}^{M} \left(h_\theta(x^{m}) - y^{m} \right) \cdot 1 \\ \frac{1}{M} \sum_{m} \left(h_\theta(x^{m}) - y^{m} \right) \cdot x_1^{m} \\ \frac{1}{M} \sum_{m} \left(h_\theta(x^{m}) - y^{m} \right) \cdot x_2^{m} \\ \vdots \\ \frac{1}{M} \sum_{m} \left(h_\theta(x^{m}) - y^{m} \right) \cdot x_n^{m} \end{bmatrix}$$

Q5]	Show: Hyperbolic Set $\{x \in \mathbb{R}^2_+ \mid x_1 x_2 \geq 1\}$ is Convex

Consider two points (x_1, x_2) and (y_1, y_2). Let z be a convex combination of those two points.

If $x \geq y$, then $z = \theta x + (1-\theta) y \geq y$
and $z_1 z_2 \geq y_1 y_2 \geq 1$

Suppose $y \not\geq 0$ and $x \not\geq y$ i.e. $(y_1 - x_1)(y_2 - x_2) < 0$, then:

$$
\begin{aligned}
(\theta x_1 + (1-\theta) y_1)(\theta x_2 + (1-\theta) y_2) \\
&= \theta^2 x_1 x_2 + (1-\theta)^2 y_1 y_2 + \theta(1-\theta) x_1 y_2 + \theta(1-\theta) x_2 y_1 \\
&= \theta x_1 x_2 + (1-\theta) y_1 y_2 - \theta(1-\theta)(y_1 - x_1)(y_2 - x_2) \\
&\geq 1
\end{aligned}
$$

Show: $\{x \in \mathbb{R}^n_+ \mid \prod_{i=1}^{n} x_i \geq 1\}$ is convex

If $\prod_i x_i \geq 1$ and $\prod_i y_i \geq 1$, then

$$\prod_i (\theta x_i + (1-\theta) y_i) \geq \prod x_i^{\theta} y_i^{1-\theta} = \left(\prod_i x_i\right)^{\theta} \left(\prod_i y_i\right)^{1-\theta} \geq 1$$

Hence, it is convex.

Q6] Show whether following funct^n are Convex, concave, both, or none

a) $f(x) = e^x - 1$ on $\mathbb{R}$

Here, e^x is a convex funct^n. $f(x)$ is shifted by one unit.
$\therefore$ It is convex

b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$

$$\nabla f(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

It is neither positive semidefinite nor negative semidefinite.
$\therefore$ f is neither convex nor concave.

c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$

$$\nabla f(x) = \begin{bmatrix} \frac{-1}{x_1^2 x_2} \\ \frac{-1}{x_1 x_2^2} \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix}$$

$$\Rightarrow \nabla^2 f(x) = \frac{1}{x_1 x_2} \begin{bmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{bmatrix} \succeq 0$$

$\therefore$ f is convex

d) $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$

$$\nabla f(x) = \begin{bmatrix} \frac{1}{x_2} \\ \frac{-x_1}{x_2^2} \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 0 & \frac{-1}{x_2^2} \\ \frac{-1}{x_2^2} & \frac{2x_1}{x_2^3} \end{bmatrix}$$

$\nabla^2 f(x)$ is neither positive nor negative semidefinite

$\therefore f$ is neither convex nor concave.

e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R} \times \mathbb{R}_{++}$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{2}{x_2} & \frac{-2x_1}{x_2^2} \\ \frac{-2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{bmatrix}$$

$$= \left(\frac{2}{x_2}\right) \begin{bmatrix} 1 \\ \frac{-2x_1}{x_2} \end{bmatrix} \begin{bmatrix} 1 & \frac{-2x_1}{x_2} \end{bmatrix} \succeq 0$$

$\therefore f$ is convex.

f) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$ where $0 \leq \alpha \leq 1$ on $\mathbb{R}^2_{++}$

$$\nabla^2 f(x) = \begin{bmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & (1-\alpha)(-\alpha) x_1^{\alpha} x_2^{-\alpha-1} \end{bmatrix}$$

$$= \alpha(1-\alpha) x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} \frac{-1}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{-1}{x_2^2} \end{bmatrix}$$

$$= -\alpha(1-\alpha)\, x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} \frac{1}{x_1} \\ \frac{-1}{x_2} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} \\ \frac{-1}{x_2} \end{bmatrix}^T \preceq 0$$

$\therefore$ f is concave.

Q7] $$f(\theta) = -y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

where $$h_\theta(x) = \frac{1}{1+e^{(-\theta^T x)}}$$

Prove that $f(\theta)$ is a convex function of the vector $\theta \in \mathbb{R}^n$ for each $x \in \mathbb{R}^n$ and $y \in [0,1]$.

Here, $$f(\theta) = -y\log(h_\theta(x)) - \log(1-h_\theta(x)) + y\log(1-h_\theta(x))$$

Rearranging $$= -\log(1-h_\theta(x)) + y\log(1-h_\theta(x)) - y\log(h_\theta(x))$$

$$= -\log(1-h_\theta(x)) + y.\log\left(\frac{1-h_\theta(x)}{h_\theta(x)}\right)$$

Substituting value of $h_\theta(x)$ in second term, we get

$$f(\theta) = -\log(1-h_\theta(x)) + y.\log\left(\frac{1-\frac{1}{1+e^{-\theta^T x}}}{\frac{1}{1+e^{-\theta^T x}}}\right)$$

$$= -\log(1-h_\theta(x)) + y.\log\left(\frac{\cancel{1}+e^{-\theta^T x}-\cancel{1}}{\cancel{1+e^{-\theta^T x}}} \times \frac{\cancel{1+e^{-\theta^T x}}}{1}\right)$$

$$= -\log(1-h_\theta(x)) + y.\log(e^{-\theta^T x})$$

$$= -\log(1-h_\theta(x)) + (-y\theta^T x)$$

$$f(\theta) = \quad A \quad + \quad B$$

Here, B is affine transformation of θ, where $y \in [0,1]$

$\therefore$ B is convex ——— ①

For A,

$$A = -\log(1 - h_\theta(x))$$

$$= -\log\left(1 - \frac{1}{1+e^{-\theta^T x}}\right)$$

$$= -\log\left(\frac{\cancel{1}+e^{-\theta^T x}\cancel{-1}}{1+e^{-\theta^T x}}\right)$$

$$= -\log\left(\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right) = \cancel{\log\left(\frac{1+e^{-\theta^T x}}{e^{-\theta^T x}}\right)}$$

$$= -\log e^{-\theta^T x} + \log(1+e^{-\theta^T x})$$

$$= \theta^T x + \log(1+e^{-\theta^T x})$$

Now, $$\frac{\partial A}{\partial \theta} = x + \frac{1}{1+e^{-\theta^T x}} \cdot e^{-\theta^T x} \cdot (-x)$$

$$= x - \frac{x e^{-\theta^T x}}{1+e^{-\theta^T x}} = x\left(\frac{1+\cancel{e^{-\theta^T x}} - \cancel{e^{-\theta^T x}}}{1+e^{-\theta^T x}}\right)$$

$$\frac{\partial A}{\partial \theta} = \frac{x}{1+e^{-\theta^T x}}$$

$$\frac{\partial^2 A}{\partial \theta^2} = x \cdot \frac{-1}{(1+e^{-\theta^T x})^2} \cdot (e^{-\theta^T x})(-x)$$

$$\frac{\partial^2 A}{\partial \theta^2} = \frac{e^{-\theta^T x}}{(1+e^{-\theta^T x})^2} \cdot x^2$$

Here, $\frac{\partial^2 A}{\partial \theta^2} \geq 0$

$\therefore$ A is convex ——— (2)

from eq^n (1) & (2), it is proved that $f(\theta)$ is a convex function.

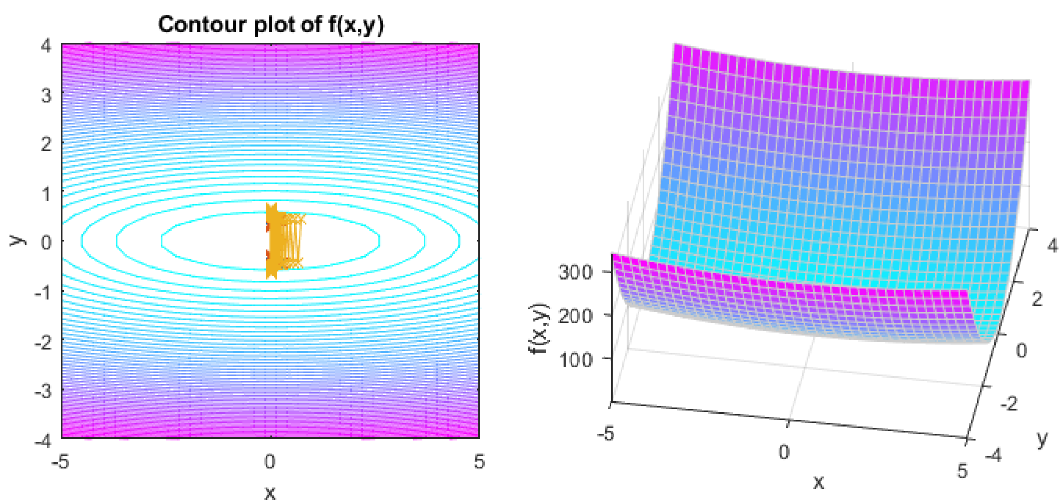
Question 8:

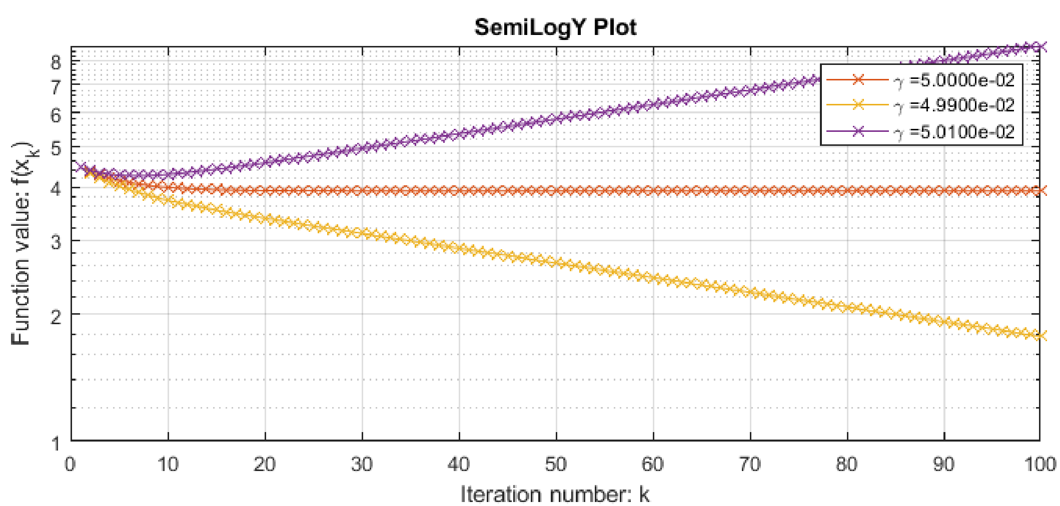
a) The hessian of the function provides a value of 2 as the least eigen value(mu) and 40 as the highest eigenvalue(L).
b) The interval of (A,B) is set to (0,2/L)
c)

EXERCISE: Enter the step sizes (values of gamma) to be compared. Choose epsn < 1e-3 if required.

```
epsn = 1e-4;

gammaVec = [2/40, (2/40) - epsn, (2/40) + (epsn)];
ng = numel(gammaVec);
```





The red line belongs to the (B+Epsilon), purple line belongs to (B) and yellow line belongs to (B-epsilon)

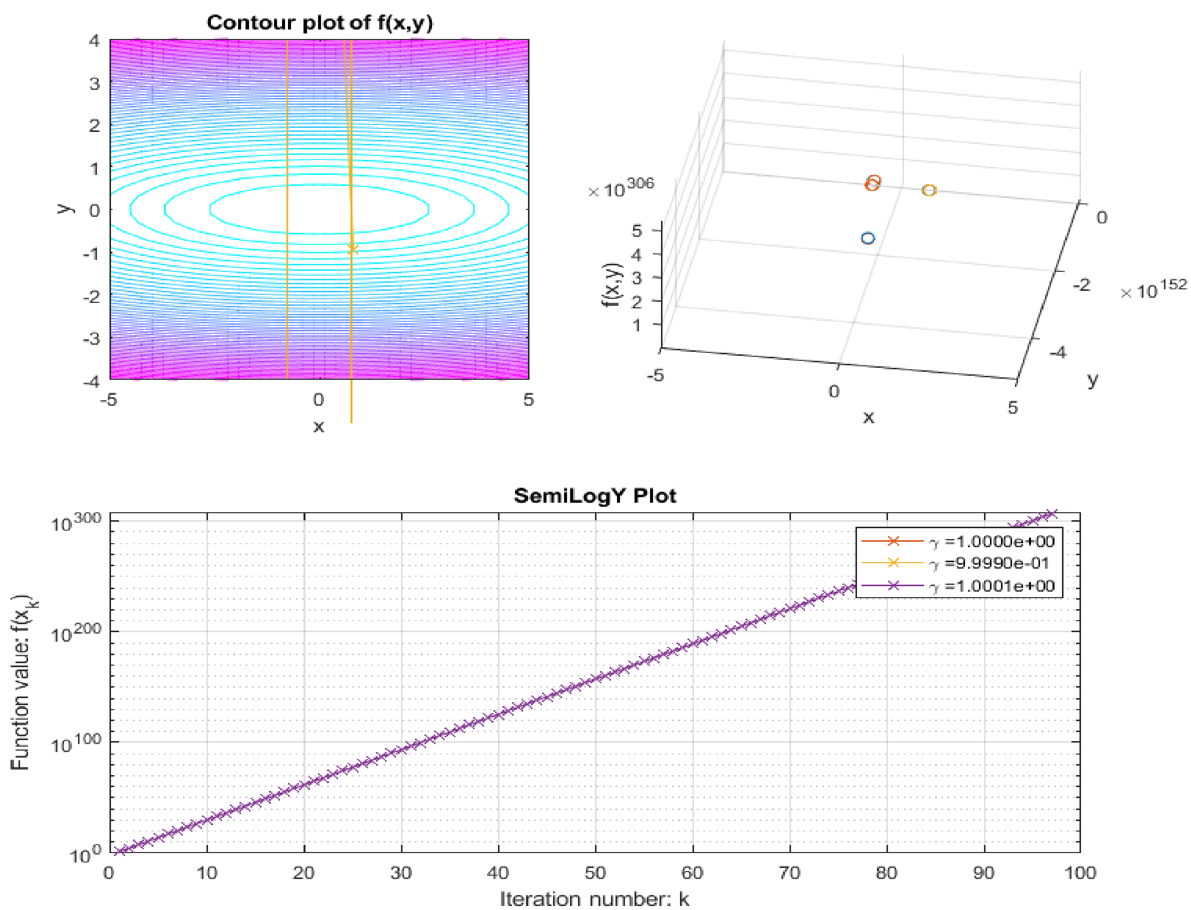
From the above graphs, it can be observed that when the step size was increased marginally than lemma growth interval, the function values do not converge to the minimum.

d)

EXERCISE: Enter the step sizes (values of gamma) to be compared. Choose epsn < 1e-3 if required.

```
epsn = 1e-4;

gammaVec = [1, (1) - epsn, (1) + (epsn)];
ng = numel(gammaVec);
```

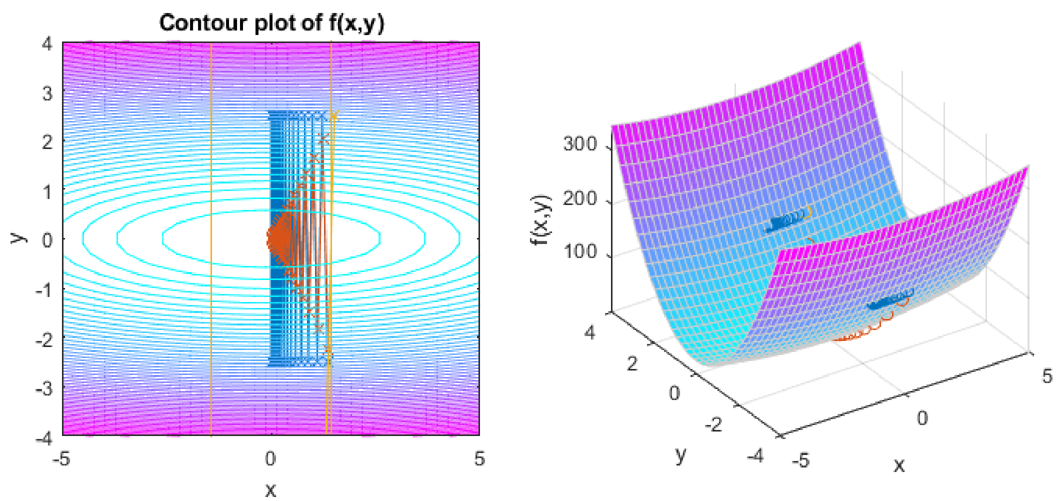


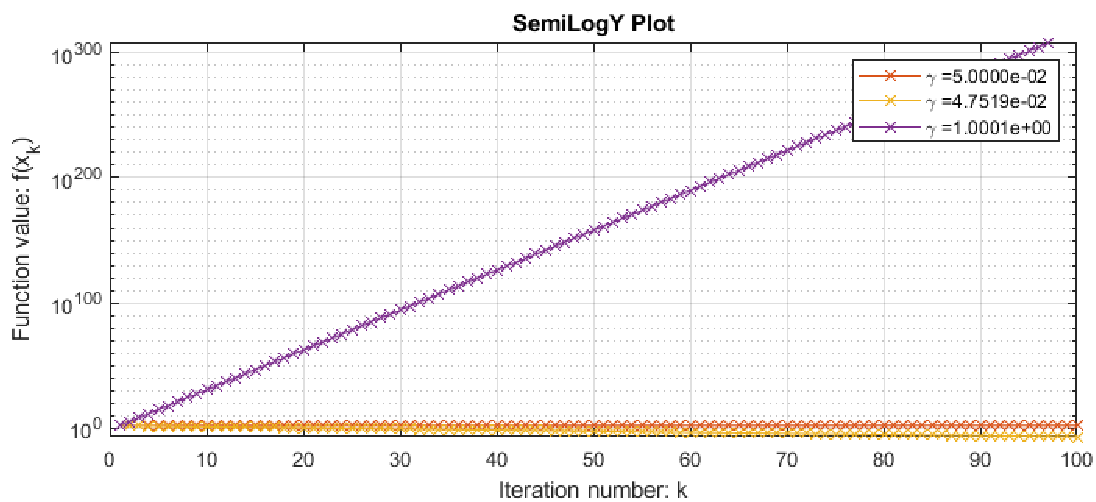
e)

EXERCISE: Enter the step sizes (values of gamma) to be compared. Choose epsn < 1e-3 if required.

```
epsn = 1e-4;

gammaVec = [2/40, (2/42) - epsn, 1 + (epsn)];
ng = numel(gammaVec);
```





The function has a linear increase for a gamma of 1.