Amir Hossein Eyvazkhani

1747696

exo2

Task 1)

a) $\nabla f = \begin{pmatrix} df/dx \\ df/dy \\ df/dz \end{pmatrix} = \begin{pmatrix} z^2 e^y \cos y \\ \cos y (xz^2 e^y) + xz^2 e^y(-\sin y) \\ 2zxe^y \cos y \end{pmatrix} = \begin{pmatrix} z^2 e^y \cos y \\ xz^2 e^y (\cos y - \sin y) \\ 2zxe^y \cos y \end{pmatrix}$

b) $\nabla g = \begin{pmatrix} \frac{dg}{dx} \\ \frac{dg}{dy} \\ \frac{dg}{dz} \end{pmatrix} = \begin{pmatrix} \frac{1}{x^2+y^2+z^2}(2x) \\ \frac{1}{x^2+y^2+z^2}(2y) \\ \frac{1}{x^2+y^2+z^2}(2z) \end{pmatrix}$

Task 2)

a) $\frac{df}{dx} = -2(a-x) + 2b(y-x^2)(-2x) = -2a + 2x - 4bxy + 4bx^3$

$\frac{df}{dy} = 2b(y-x^2) = 2by - 2bx^2$

$\nabla f = \begin{pmatrix} df/dx \\ df/dy \end{pmatrix} = \begin{pmatrix} 4bx^3 + 2x(1-2by) - 2a \\ 2by - 2bx^2 \end{pmatrix}$

b) $\nabla f = 0 \Rightarrow \begin{cases} 4bx^3 + 2x(1-2by) - 2a = 0 \overset{\mathrm{I}}{\Rightarrow} 2x - 2a = 0 \Rightarrow x = a \ \mathrm{II} \\ 2by - 2bx^2 = 0 \Rightarrow y - x^2 = 0 \Rightarrow y = x^2 \ \mathrm{I} \quad \Rightarrow y = a^2 \end{cases}$

$f^*(x,y) = (a-a)^2 + b(a^2-a^2)^2 = 0 \quad \Leftarrow (x,y)^* = (a, a^2)$

$\frac{df}{d^2x} = 12bx^2 + 2(1-2by) \qquad \frac{df}{dxdy} = -4bx$

$\frac{df}{dydx} = -4bx \qquad \frac{df}{d^2y} = 2b$

$\Rightarrow H_f = \begin{pmatrix} 12bx^2 - 4by + 2 & -4bx \\ -4bx & 2b \end{pmatrix}$

$\det(H) > 0 \xrightarrow{d^{1\times 1} > 0} 12bx^2 - 4by + 2 > 0 \xRightarrow{(a,a^2)} 12ba^2 - 4ba + 2 > 0$

$\Rightarrow b(12a^2 - 4a) > -2 \Rightarrow b > \dfrac{-2}{12a^2-4a} \ : a > 1$

$\quad\quad b < \dfrac{+2}{12a^2-4a} \ ; 0 < a < 1$ II

$d^{2\times 2} > 0$

$\to 24b^2x^2 - 8b^2y + 4b - 16b^2x^2 > 0 \Rightarrow \cancel{8b^2x^2} \cancel{- 8b^2y} + 4b > 0$

$\xRightarrow{(a,a^2)} 4b > 0 \Rightarrow b > 0$ ✓

$\text{trace}(H) > 0 \Rightarrow 12bx^2 - 4by + 2 + 2b > 0 \xRightarrow{(a,a^2)} 12ba^2 - 4ba + 2 + 2b > 0$

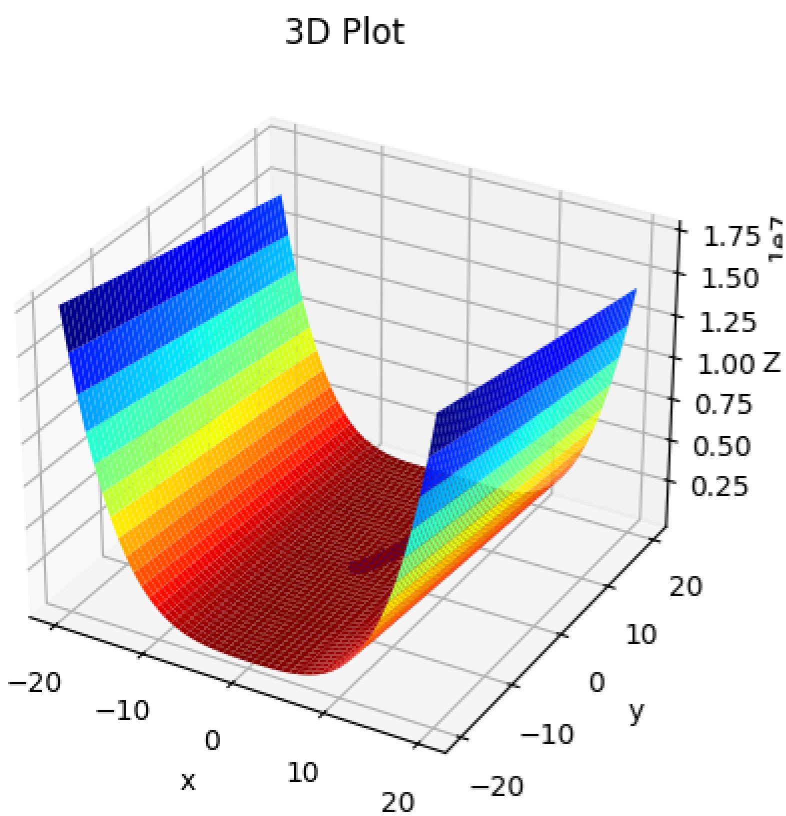
$\Rightarrow b(\underbrace{6a^2 - 2a + 1}_{>0}) > -1 \Rightarrow b > \dfrac{-1}{6a^2-2a+1}$ I

I, II $\to$ It is positive definite only if I, II hold.

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
```

In [ ]:
```
lr = 0.005 # Learning rate
epochs = 1000
eps = 0.0001

def f(x, y, a, b):
    return (a - x)**2 + b*(y - x**2)**2

def gradient_f_x(x, y, a, b):

    return 4*b*x**3 + 2*x - 4*b*x*y -2*a

def gradient_f_y(x, y, a, b):
    return 2*b*y - 2*b*x**2

def draw_progress(Xs, Ys, a, b):
    fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
    reso = 50
    X = np.linspace(-20, 20, reso)
    Y = np.linspace(-20, 20, reso)
    X, Y = np.meshgrid(X, Y)
    Z = (a - X)**2 + b*(Y - X**2)**2
    Zs = (a - Xs)**2 + b*(Ys - Xs**2)**2
    surf = ax.plot_surface(X, Y, Z, cmap='jet_r',
        linewidth=0, antialiased=True, label="search space")
    ax.set_xlabel("x")
    ax.set_ylabel("y")
    ax.set_zlabel('Z', labelpad=1)
    ax.set_title('3D Plot')
    ax.plot(
        Xs, Ys, Zs, '.-b',
        label = 'gradient descent'
    )
    plt.show()

def gradient_decent_booth(x,y, a, b):
    value_before = f(x, y, a, b)
    Xs = [x]
    Ys = [y]
    for i in range(epochs):
        x = x - lr*gradient_f_x(x,y, a, b)
        y = y - lr*gradient_f_y(x,y, a, b)
        x = np.clip(x, -20, 20)
        y = np.clip(y, -20, 20)
        value = f(x, y, a, b)
        Xs.append(x)
        Ys.append(y)
        if (abs(
            value_before - value
        ) < eps):
            draw_progress(np.array(Xs), np.array(Ys), a, b)
            print(f'final x : {x} - final y : {y} - final f(x, y) : {value_before}'
```

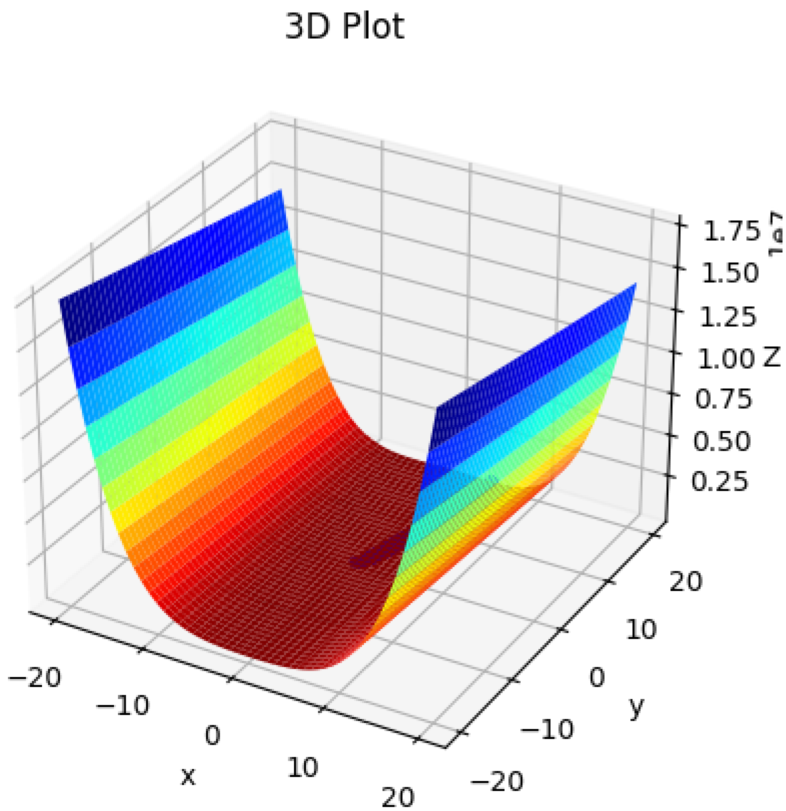
```
            return (x, y)
        value_before = value
    draw_progress(np.array(Xs), np.array(Ys), a, b)
    print(f'final x : {x} - final y : {y} - final f(x, y) : {value_before}')

gradient_decent_booth(0, 0, 3, 100)
```




```
final x : 2.9300666980564105 - final y : 8.585290855059196 - final f(x, y) : 0.00498
9967065332935
```

Out[ ]: (2.9300666980564105, 8.585290855059196)

In [ ]:
```
gradient_decent_booth(0, 0, 10, 100)
```




```
final x : 3.033030927711535 - final y : 9.199276608454694 - final f(x, y) : 48.53865
8054224
```

In [ ]:

Task 4) $L(w) = \frac{1}{n} \sum_{i=1}^{n} l(f_w(x_i), y_i)$ I

$$l(f_w(x_i), y_i) = -\log P(y = y_i / x_i ; w)$$

$$= \sum_{j=1}^{k} \left[-\log \left(\frac{e^{f_w(x_i)_j}}{\sum_{l=1}^{k} e^{f_w(x_i)_j}} \right) \right] \delta_{y_i c_j}$$

for each $1 \leq m \leq k$ $\Longrightarrow l(f_w(x_i)_j, y_i) = -\delta_{y_i c_m} \log \frac{e^{f_w(x_i)_m}}{\sum_{l=1}^{k} e^{f_w(x_i)_m}}$ II

$f_w(x) = Wx$ III

$$\xrightarrow{\text{I}, \text{II}, \text{III}} L(w_m) = \frac{-1}{n} \sum_{i=1}^{n} \delta_{y_i c_m} \log \frac{e^{w_m x_i}}{\sum_{l=1}^{k} e^{w_l x_i}}$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \delta_{y_i c_m} \left[w_m x_i \right] - \left(\log \sum_{l=1}^{k} e^{w_l x_i} \right)$$

$$\Rightarrow \frac{dL(w_m)}{dw_m} = \frac{-1}{n} \cdot \sum_{i=1}^{n} \delta_{y_i c_m} \cdot x_i - \frac{(x_i) e^{w_m x_i}}{\sum_{l=1}^{k} e^{w_l x_i}}$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left(\delta_{y_i c_m} - \frac{e^{w_m x_i}}{\sum_{l=1}^{k} e^{w_l x_i}} \right) x_i$$