AFRICAN INSTITUTE FOR MATHEMATICAL SCIENCES

(AIMS RWANDA, KIGALI)

Name: Emmanuel Ansah	Assignment Number: 1
Course: Data-Driven Optimization	Date: February 2, 2025

1.

Show the following,

1a) For the sigmoid function $\sigma(t) = \frac{1}{1+e^{-t}}$, show that its derivative is $\sigma'(t) = \sigma(t)[1 - \sigma(t)]$.

1b) If $g : \mathbb{R}^n \to \mathbb{R}$ and $g(x) = \log(\sigma(f(x)))$, then show that

$$\nabla g(x) = [1 - \sigma(f(x))]\nabla f(x).$$

1c* (left as an exercise to the curious) Show that the Hessian matrix of $g(x) = \log(\sigma(f(x)))$ (defined in 1b)) is given by

$$[1 - \sigma(f(x))]\nabla^2 f(x) - \sigma(f(x))\nabla f(x)\nabla f(x)^T.$$

Solution

1a)

Given the function,

$$\sigma(t) = \frac{1}{1+e^{-t}},$$

$$\sigma(t) = (1+e^{-t})^{-1}.$$

Differentiating with respect to t ,

$$\sigma'(t) = -\frac{d}{dt}(1+e^{-t})^{-1}.$$

$$\sigma'(t) = -(-e^{-t})(1+e^{-t})^{-2}$$

$$= e^{-t}(1+e^{-t})^{-2}.$$

Rewriting as,

$$\sigma'(t) = \frac{(1+e^{-t}) - 1}{(1+e^{-t})^2}.$$

splitting,

$$\sigma'(t) = \frac{1+e^{-t}}{(1+e^{-t})^2} - \frac{1}{(1+e^{-t})^2}.$$

$$\sigma'(t) = \frac{1}{1+e^{-t}} - \frac{1}{(1+e^{-t})^2}.$$

Factoring out $\sigma(t)$,

$$\sigma'(t) = \sigma(t)\left(1 - \frac{1}{1+e^{-t}}\right).$$

Since $\frac{1}{1+e^{-t}} = \sigma(t)$,

$$\therefore \quad \sigma'(t) = \sigma(t)[1-\sigma(t)].$$

1b)

Given,

$$g(x) = \log(\sigma(f(x))),$$

using the definition of σ,

$$g(x) = \log\left(\frac{1}{1+e^{-f(x)}}\right).$$

Simplifying,

$$g(x) = -\log(1+e^{-f(x)}).$$

Differentiating with respect to x,

$$\nabla g(x) = -\frac{d}{dx}\log(1+e^{-f(x)}).$$

$$\nabla g(x) = -\left(\frac{-e^{-f(x)}}{1+e^{-f(x)}}\right)\nabla f(x).$$

Simplifying,

$$\nabla g(x) = \frac{e^{-f(x)}}{1+e^{-f(x)}}\nabla f(x).$$

Since $\frac{e^{-f(x)}}{1+e^{-f(x)}} = 1-\sigma(f(x))$,

$$\therefore \quad \nabla g(x) = [1-\sigma(f(x))]\nabla f(x).$$

1c*)
Given,

$$g(x) = \log(\sigma(f(x))),$$

From (1b),

$$\nabla g(x) = [1 - \sigma(f(x))]\nabla f(x).$$

Differentiating with respect to x,

$$\nabla^2 g(x) = \nabla\left([1 - \sigma(f(x))]\nabla f(x)\right).$$

using product rule,

$$\nabla^2 g(x) = \nabla[1 - \sigma(f(x))] \cdot \nabla f(x)^T + [1 - \sigma(f(x))]\nabla^2 f(x).$$

Computing $\nabla[1 - \sigma(f(x))]$,

$$\nabla[1 - \sigma(f(x))] = -\nabla\sigma(f(x)).$$

From (1a),

$$\nabla\sigma(f(x)) = \sigma(f(x))[1 - \sigma(f(x))]\nabla f(x).$$

Substituting back,

$$\nabla[1 - \sigma(f(x))] = -\sigma(f(x))[1 - \sigma(f(x))]\nabla f(x).$$

Now,

$$\nabla^2 g(x) = -\sigma(f(x))[1 - \sigma(f(x))]\nabla f(x)\nabla f(x)^T + [1 - \sigma(f(x))]\nabla^2 f(x).$$

Factoring out $[1 - \sigma(f(x))]$,

$$\nabla^2 g(x) = [1 - \sigma(f(x))]\nabla^2 f(x) - \sigma(f(x))\nabla f(x)\nabla f(x)^T.$$

Thus, the Hessian,

$$\nabla^2 g(x) = [1 - \sigma(f(x))]\nabla^2 f(x) - \sigma(f(x))\nabla f(x)\nabla f(x)^T.$$

2.

Given the function

$$f(x) = 100(x_2 - x_1^2) + (1 - x_1)^2$$

2a) Use the Matplotlib to generate a 3D plot of the function.

2b) Generate a contour plot of f.

2c) Determine the gradient of $f(x)$ using Autograd under PyTorch.

2d) Determine the gradient at the point $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

2e) Is f a convex function?

Solution

2a

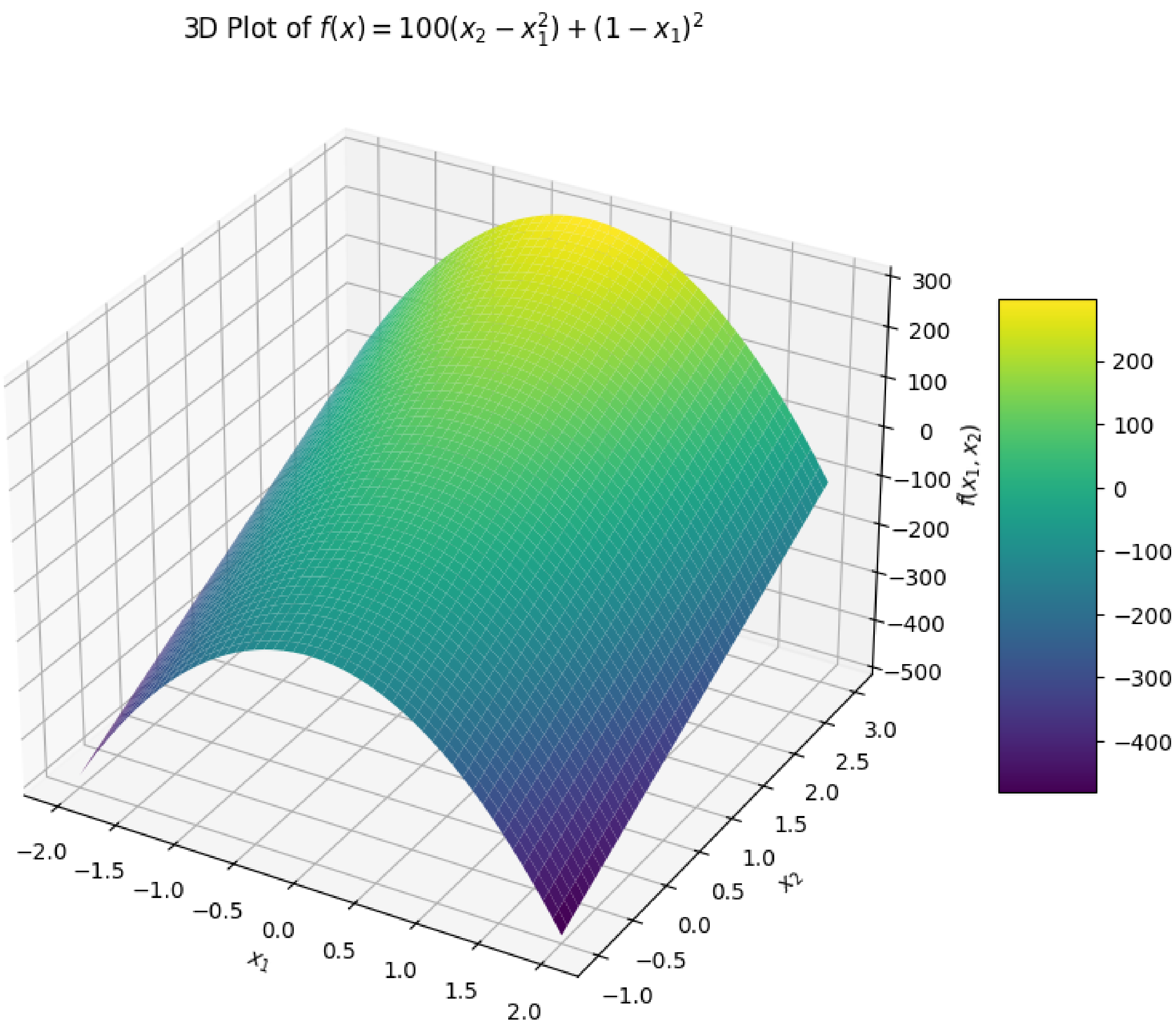

Figure 1: 3D plot of the function

2b

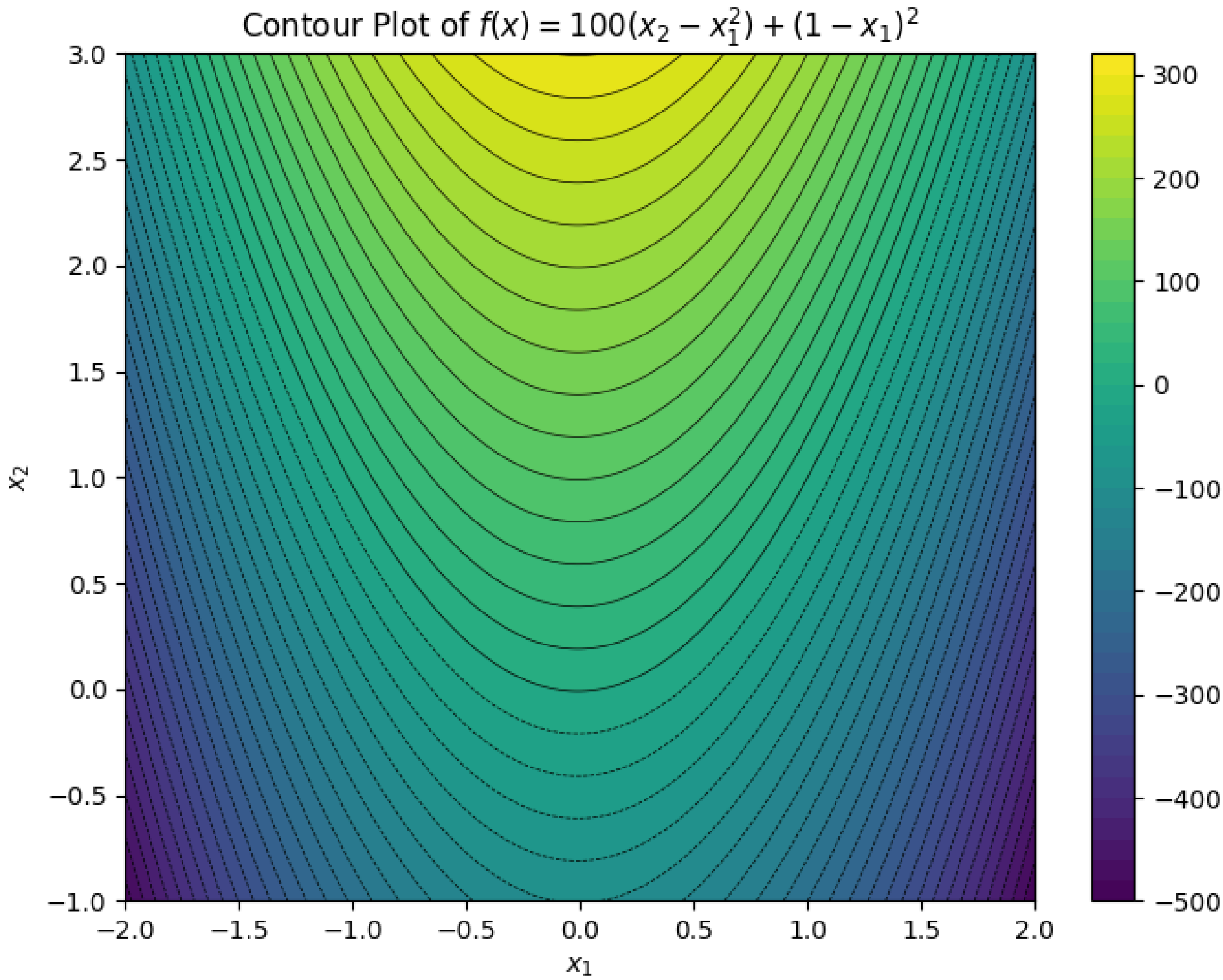

Figure 2: contour plot of the function

2d. Given,

$$f(x) = 100(x_2 - x_1^2) + (1 - x_1)^2,$$

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{\partial}{\partial x_1}\left[100(x_2 - x_1^2) + (1 - x_1)^2\right] \\
&= 100(-2x_1) - 2(1 - x_1) \\
\frac{\partial f}{\partial x_1} &= -200x_1 - 2(1 - x_1).
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial x_2} &= \frac{\partial}{\partial x_2}\left[100(x_2 - x_1^2) + (1 - x_1)^2\right] \\
\frac{\partial f}{\partial x_2} &= 100
\end{aligned}$$

At $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$,

$$\frac{\partial f}{\partial x_1} = -200(1) - 2(1-1) = -200.$$

$$\frac{\partial f}{\partial x_2} = 100.$$

The gradient of $f(x)$ is,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}.$$

At $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, the gradient is:

$$\nabla f(x) = \begin{bmatrix} -200 \\ 100 \end{bmatrix}.$$

2e.

$$\begin{aligned} f(x) &= 100(x_2 - x_1^2) + (1 - x_1)^2 \\ &= 100x_2 - 100x_1^2 + 1 - 2x_1 + x_1^2 \\ &= -99x_1^2 - 2x_1 + 100x_2 + 1 \end{aligned}$$

Finding Gradient,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = -198x_1 - 2, \quad \frac{\partial f}{\partial x_2} = 100$$

$$\Rightarrow \quad \nabla f(x) = \begin{bmatrix} -198x_1 - 2 \\ 100 \end{bmatrix}$$

Hessian Matrix,

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = -198, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 0, \quad \frac{\partial^2 f}{\partial x_2^2} = 0$$

$$H = \begin{bmatrix} -198 & 0 \\ 0 & 0 \end{bmatrix}$$

Finding Eigenvalues of H,

$$\det(H - \lambda I) = 0$$

$$\lambda_1 = -198, \quad \lambda_2 = 0$$

Since $\lambda_1 = -198 < 0$, H is not positive semi-definite.
$\therefore$ Since H is not positive semi-definite, the function $f(x)$ is not convex.