Numerical Optimization
Individual Handin 1

Dmitry Serykh (qwl888)

February 11, 2020

1 Introduction

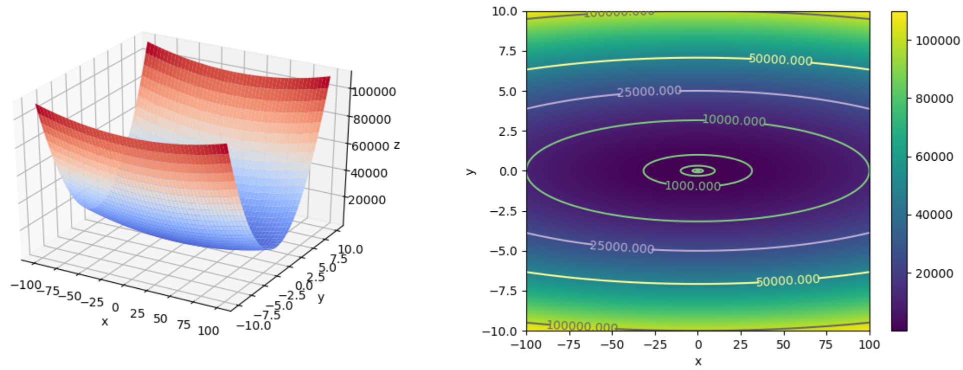
I had to retake this course due to some personal reasons. Hence, some of the solutions in this submission are based on my solutions from the previous year. Furthermore, in the report, I reduce the problems to two dimensions. However, my implementation of all, but Rosenbrock function support arbitrary values of d.

2 Ellipsoid Function

The Ellipsoid function is given by:

$$f_1(x) = \sum_{i=1}^{d} \alpha^{\frac{i-1}{d-1}} x_i^2, \alpha = 1000$$

The plots for the function can be seen on Figure 1.



(a) Surface plot

(b) Contour plot

Figure 1: Ellipsoid Function

2.1 First Derivative

I find the gradient:

$$\nabla f_1(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2\alpha x_2 \end{bmatrix}$$

2.2 Second Derivative

Then I find the Hessian matrix, but first I will consider the function:

$$f(x) = \sum_{i=1}^{N} g_i(x_i)$$

$\frac{\partial g_i(x_i)}{\partial x_j} = 0$ for $j \neq i$, since $g_i(x_i)$ is not dependent on any x_j, where $j \neq i$. And since $\frac{\partial 0}{\partial x} = 0$, for all $x \in \mathbb{R}^N$, the hessian matrix would only have nonzero values in the entries where $j = i$. Therefore we have that the Hessian is a diagonal matrix with

$$(Hf(x))_{ii} = g_i''(x_i)$$

I then use this notion in order to determine the Hessian:

$$H(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 2\alpha \end{bmatrix}$$

2.3 Finding a Minimizer

I can then solve a system of equations in order to find the stationary points:

$$\begin{bmatrix} 2x_1 \\ 2\alpha x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The system of equations has only one solution, hence $(0,0)^T$ is the only stationary point. I then determine the type of the stationary point by finding the determinant of the hessian matrix:

$$det(H(0,0)) = det\left(\begin{bmatrix} 2 & 0 \\ 0 & 2000 \end{bmatrix}\right) = 4000 > 0$$

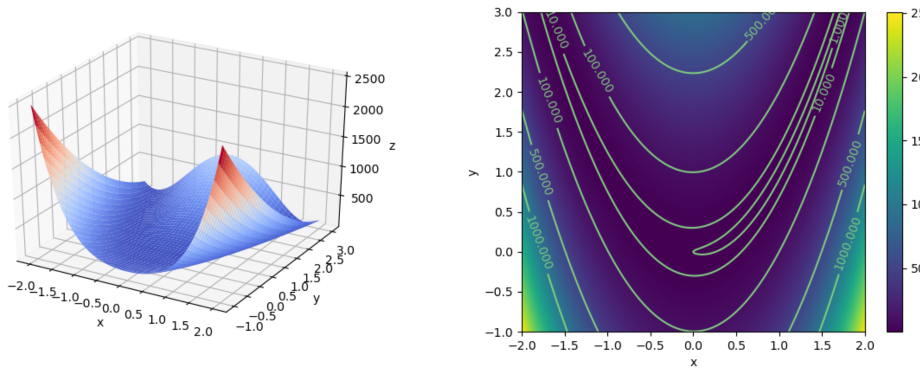
Determinant is positive, hence I conclude that $H(1,1)$ is positive definite, and $(0,0)^T$ is the only local minimizer of the Ellipsoid function.

3 Rosenbrock Function

The Rosenbrock function is given by:

$$f_2(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

The plots for the function can be seen on Figure 2.



(a) Surface plot

(b) Contour plot

Figure 2: Rosenbrock Function

3.1 First Derivative

I find the gradient by applying the chain rule:

$$\nabla f_2(x_1, x_2) = \begin{bmatrix} -400x(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

3.2 Second Derivative

Then I find the Hessian matrix:

$$H(x_1, x_2) = \begin{bmatrix} 1200x^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

3.3 Finding a Minimizer

I can then solve a system of equations in order to find the stationary points:

$$\begin{bmatrix} -400x_1(x_2 - x^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The system of equations has only one solution, hence $(1, 1)^T$ is the only stationary point. I then determine the type of the stationary point by finding the determinant of the hessian matrix:

$$det(H(1, 1)) = det\left(\begin{bmatrix} 1200 - 400 + 2 & -400 \\ -400 & 200 \end{bmatrix}\right) = 400 > 0$$

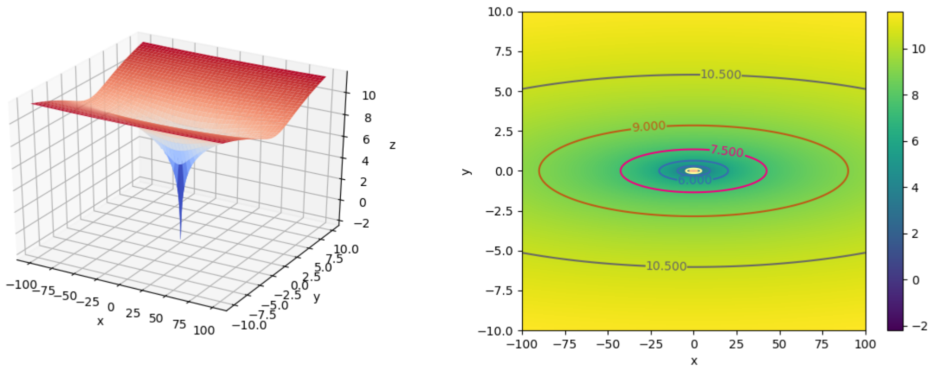
Since $\frac{df}{dx_1 dx_2} = 802 > 0$ and the determinant is positive, I conclude that $H(1, 1)$ is positive definite, and $(1, 1)^T$ is the only local minimizer of the 2D Rosenbrock function.

4 Log-Ellipsoid Function

The Log-Ellipsoid function is given by:

$$f_3(x) = \log\left(\epsilon + f_1(x)\right), \epsilon = 10^{-16}$$

The plots for the function can be seen on Figure 3.



(a) Surface plot

(b) Contour plot

Figure 3: Log-Ellipsoid Function

4.1 First Derivative

$$\nabla f_3(x_1, x_2) = \begin{bmatrix} (2x_1)/(\epsilon + x_1^2 + \alpha x_2^2) \\ (2\alpha x_2)/(\epsilon + x_1^2 + \alpha x_2^2) \end{bmatrix}$$

4.2 Second Derivative

$$H(x_1, x_2) = \begin{bmatrix} (2(\epsilon - x_1^2 + \alpha x_2^2))/(\epsilon + x_1^2 + \alpha x_2^2)^2 & -(4\alpha x_1 x_2)/(\epsilon + x_1^2 + \alpha x_2^2)^2 \\ -(4\alpha x_1 x_2)/(\epsilon + x_1^2 + \alpha x_2^2)^2 & (2\alpha(\epsilon + x_1^2 - \alpha x_2^2))/(\epsilon + x_1^2 + \alpha x_2^2)^2 \end{bmatrix}$$

5 Attractive Sector Functions

5.1 First Function

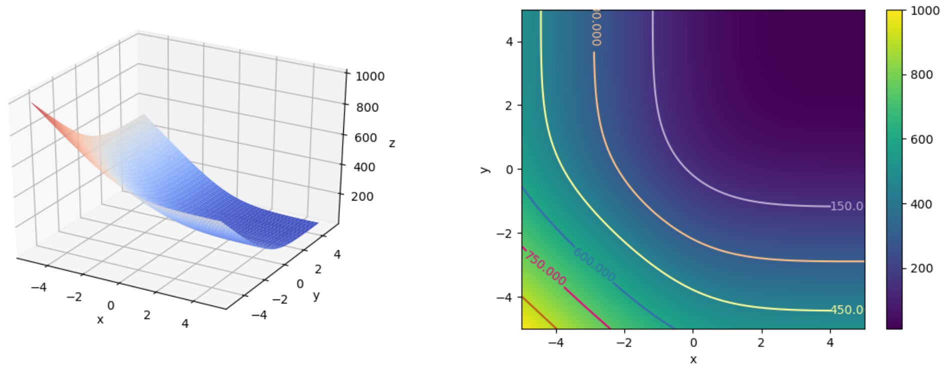
The first Attractive Sector function is given by:

$$f_4(x) = \sum_{i=1}^{d} h(x_i) + 100 \cdot h(-x_i)$$

where:

$$h(x) = \frac{\log(1 + \exp(q \cdot x))}{q}, q = 10^8$$

The plots for the function can be seen on Figure 4.



(a) Surface plot

(b) Contour plot

Figure 4: First Attractive Sector Function

5.2 First Derivative of the First Function

$$\nabla f_4(x_1, x_2) = \begin{bmatrix} (100 + e^{qx_1})/(1 + e^{qx_1}) \\ (100 + e^{qx_2})/(1 + e^{qx_2}) \end{bmatrix}$$

5.3 Second Derivative of the First Function

This function is similar to the Ellipsoid function, since it is also a sum of independent functions. Hence, we can use the property that we have shown in Section 2.

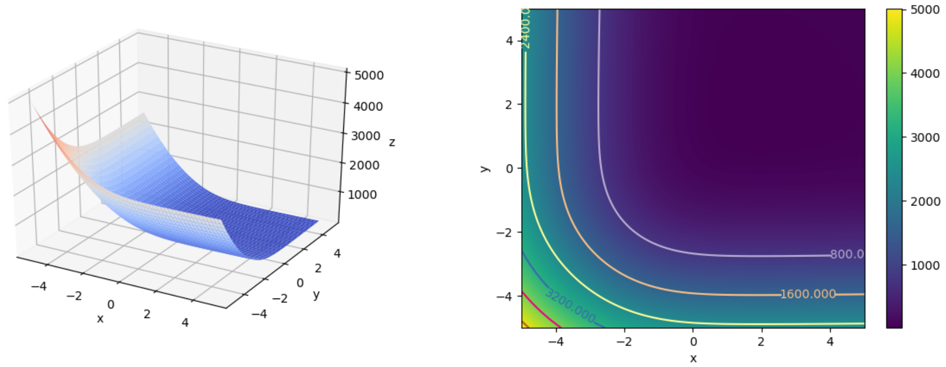
$$H(x_1, x_2) = \begin{bmatrix} -(99e^{qx_1}q)/(1 + e^{qx_1})^2 & 0 \\ 0 & -(99e^{qx_2}q)/(1 + e^{qx_2})^2 \end{bmatrix}$$

5.4 Second Function

The second Attractive Sector function is then given by:

$$f_4(x) = \sum_{i=1}^{d} h(x_i) + 100 \cdot h(-x_i)$$

where $h(x)$ is the same as in the first version. The plots for the function can be seen on Figure 5.



(a) Surface plot

(b) Contour plot

Figure 5: Second Attractive Sector Function

5.5 Overflow in the Exponential Function

Computation of h-function can be troublesome, since the exponential function can overflow. However this can be mitigated by replacing the function with an equivalent. I will prove that following property holds using proof by case:

$$\log(1 + \exp(x)) = \log(1 + \exp(-|x|)) + \max(x, 0)$$

There are two cases:

- $x \leq 0$. This case is trivial

$$\begin{aligned} \log(1 + \exp(-|x|)) + \max(x, 0) &= \log(1 + \exp(x)) + 0 \\ &= \log(1 + \exp(x)) \end{aligned}$$

- $x > 0$

$$\begin{aligned} \log(1 + \exp(-|x|)) + \max(x, 0) &= \log(1 + \exp(-x)) + x \\ &= \log\left(1 + \frac{1}{e^x}\right) + x \\ &= \log\left(\frac{1 + e^x}{e^x}\right) + x \\ &= \log(1 + e^x) - \log\left(e^x\right) + x \\ &= \log(1 + e^x) \end{aligned}$$

■

This formulation is beneficial in a context of a computer implementation, since when x is a large number, such as 10^8 in the Attractive-Sector function. The value of e^x would be a huge number that is much larger than 10^{32} or even 10^{64}. Therefore, it can not fit in any type of CPU registers and would result in a overflow. $-|x| > 0$ for all real values of x, hence $e^{-|x|}$ would become a very small number for high values of x and would be rounded to zero, which would eliminate the overflow problem.