Numerical Optimization
Week 1

Xiaohan Wang gkt918

11 February

1 Brief introduction

In the first week we reviewed some basics of Linear Algebra, Conditioning and Calculus, and implemented some basic functions using python.

2 Programming

2.1 Function 1

I have plotted the image of the f1 function in 2D and 3D with the dimension of x being 2 ($-0.1 \leq x_1, x_2 \leq 0.1$).

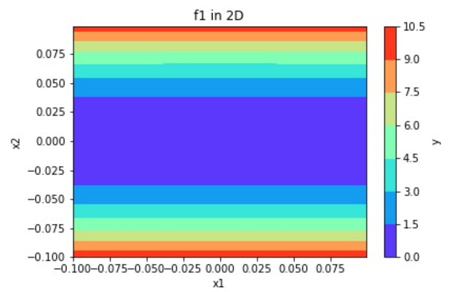

Figure 1: f1 in 2D

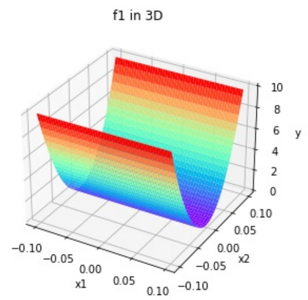

Figure 2: f1 in 3D

Its gradient is $2 * \alpha^{\frac{i-1}{d-1}} x_i$ for every x_i and Hessian matrix is

$$H(f_1) = \bigtriangledown^2 f_1(x) = \begin{bmatrix} 2 * \alpha^{\frac{0}{d-1}} & 0 & \dots & 0 \\ 0 & 2 * \alpha^{\frac{1}{d-1}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 * \alpha^{\frac{d-1}{d-1}} \end{bmatrix}$$

I picked 10 points at random and calculated their function values, first and second derivatives through an online calculator. At the same time, I also take these points as input to my implementation and record the value of the output. After comparison, the function value, the first derivative and the second derivative of each point are the same. So it proves that the function is implemented correctly.

2.2 Function 2

I have plotted the image of the f2 function in 2D and 3D ($-0.1 \leq x_1, x_2 \leq 0.1$).

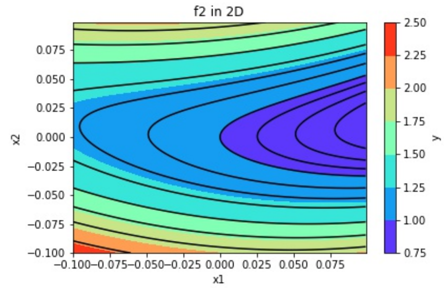

Figure 3: f2 in 2D

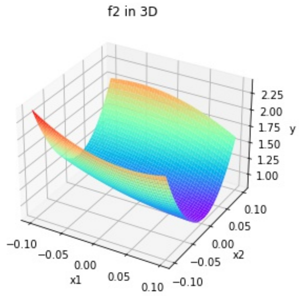

Figure 4: f2 in 3D

Its gradient is $(2x_1 - 2 + 400x_1^3 - 400x_1x_2, 200x_2 - 200x_2^2)$ and Hessian matrix is

$$H(f_2) = \bigtriangledown^2 f_2(x) = \begin{bmatrix} 2 + 1200x_1^2 - 400x_2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

The test method is the same as function 1.

2.3 Function 3

I have plotted the image of the f3 function in 2D and 3D ($-0.1 \leq x_1, x_2 \leq 0.1$).

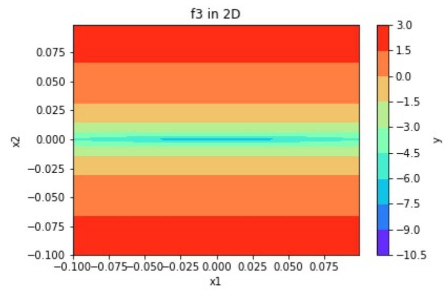

Figure 5: f3 in 2D

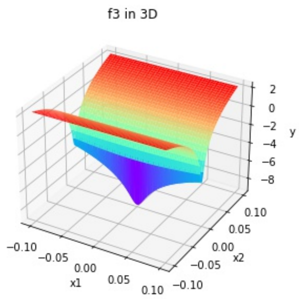

Figure 6: f3 in 3D

Its gradient is *#todo* and Hessian matrix is

$$H(f_3) = \bigtriangledown^2 f_3(x) = [\#todo]$$

The test method is the same as function 1.

2.4 Function 4

I have plotted the image of the f4 function in 2D and 3D ($-0.1 \leq x_1, x_2 \leq 0.1$).

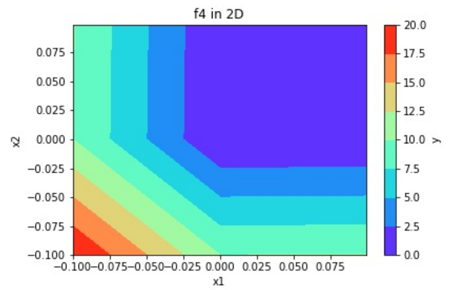

Figure 7: f4 in 2D

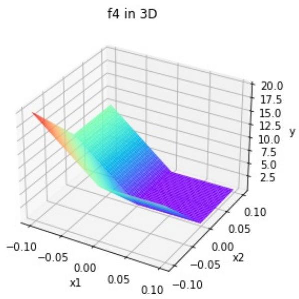

Figure 8: f4 in 3D

Its gradient is *#todo* and Hessian matrix is

$$H(f_4) = \bigtriangledown^2 f_4(x) = [\#todo]$$

The test method is the same as function 1.

2.5 Function 5

I have plotted the image of the f4 function in 2D and 3D ($-0.1 \leq x_1, x_2 \leq 0.1$).

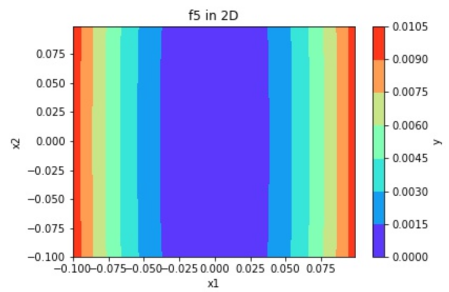

Figure 9: f5 in 2D

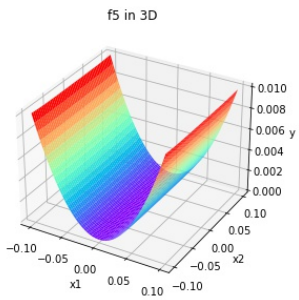

Figure 10: f5 in 3D

Its gradient is *#todo* and Hessian matrix is

$$H(f_5) = \bigtriangledown^2 f_5(x) = [\#todo]$$

The test method is the same as function 1.

3 Theoretical Exercises

3.1 Question 1

We know the Hessian of f is :

$$H(f) = \bigtriangledown^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \dots & \frac{\partial^2 f}{\partial x_1 x_N} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N x_1} & \frac{\partial^2 f}{\partial x_N x_2} & \dots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

Besides, $f(x) = \sum_{i=1}^{N} g_i(x_i)$, $x \in \mathbb{R}^N$. And it is easy to prove that for $i \neq j$, $\frac{\partial}{\partial x_j} g_i(x_i) = 0$, $\frac{\partial}{\partial x_j} g_i^{'}(x_i) = 0$. Therefore we have

$$\begin{aligned} \frac{\partial^2 f}{\partial x_i x_j} &= \frac{\partial}{\partial x_j}[\frac{\partial}{\partial x_i}[g_1(x_1) + g_2(x_2) + ... + g_N(x_N)]] \\ &= \frac{\partial}{\partial x_j}[\frac{\partial}{\partial x_i} g_1(x_1) + \frac{\partial}{\partial x_i} g_2(x_2) + ... + \frac{\partial}{\partial x_i} g_N(x_N)] \\ &= \frac{\partial}{\partial x_j}[0 + 0 + \dots + g_i^{'}(x_i)] = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x_i^2} &= \frac{\partial}{\partial x_i}[\frac{\partial}{\partial x_i}[g_1(x_1) + g_2(x_2) + ... + g_N(x_N)]] \\ &= \frac{\partial}{\partial x_i}[\frac{\partial}{\partial x_i} g_1(x_1) + \frac{\partial}{\partial x_i} g_2(x_2) + ... + \frac{\partial}{\partial x_i} g_N(x_N)] \\ &= \frac{\partial}{\partial x_i}[0 + 0 + \dots + g_i^{'}(x_i)] = g_i^{''}(x_i) \end{aligned}$$

Now, we have

$$H(f) = \bigtriangledown^2 f(x) = \begin{bmatrix} g_1^{''}(x_1) & 0 & \dots & 0 \\ 0 & g_2^{''}(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_N^{''}(x_N) \end{bmatrix}$$

It means that $(\bigtriangledown^2 f(x))_{ii} = g_i^{''}(x_i)$.

3.2 Question 3

At $x = (1, 1)$ point, $f_3(x)^{'} = 0$ and $f_3(x)^{''} < 0$. Therefore, point (1,1) is a minimizer.

3.3 Question 4

From the question, we know that we should prove log(1 + exp(x)) = log(1 + exp(-—x—)) + max(x, 0), which can be rewritten as

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

When $x \leq 0$, we have $-|x| = x$, $max(x, 0) = 0$, then

$$
\begin{aligned}
& log(1+e^x) - log(1+e^{-|x|}) - max(x,0) \\
= & log\frac{1+e^x}{1+e^x} = log1 = 0
\end{aligned}
$$

When $x > 0$, we have $-|x| = -x$, $max(x, 0) = x$, then

$$
\begin{aligned}
& log(1+e^x) - log(1+e^{-|x|}) - max(x,0) \\
= & log\frac{1+e^x}{1+e^{-x}} - x = log\frac{1+e^x}{1+e^{-x}} - log(e^x) \\
= & log\frac{1+e^x}{e^x(1+e^{-x})} = log\frac{1+e^x}{1+e^x} = log1 = 0
\end{aligned}
$$

Therefore, we have

$$
log(1+e^x) = log(1+e^{-|x|}) + max(x,0)
$$

Otherwise, if we use $log(1+e^x)$ to calculate the floating-point in f4, When $x \leq 0$ they have no difference. But when $x > 0$, it is easy to lose precision, because e^x enlarges x to a large number, and then reduces it to another number through log. In this process, the precision will be inaccurate due to multiple operations.
On the other hand, $log(1 + e^{-|x|}) + max(x, 0)$ can maintain the accuracy better because the x is taken out separately.
At the same time, when x is very large, e^x may be too large to calculate, while $e^{-|x|}$ will not.

4 Conclusion

From this assignment, I have a certain grasp of the basic theory of the course, and at the same time, I have completed the programming of functions, laying a foundation for future study.