Exercise 5

s192230 Manxi Lin

October 3, 2020

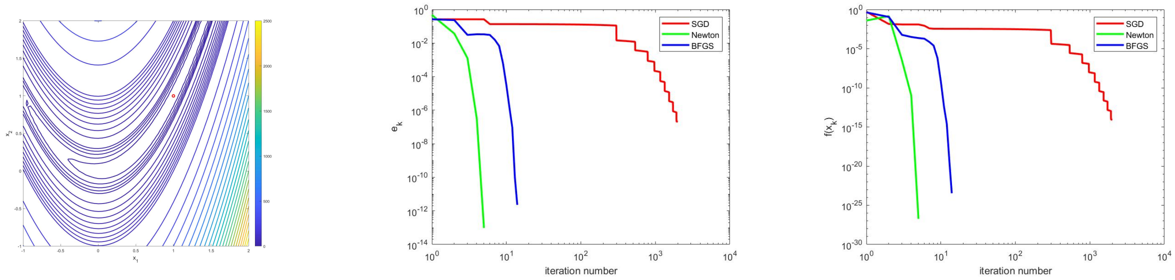
0.0.1 Exercise 5.1

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

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

0.0.2 Exercise 5.4

The answer is shown in figure 1(a). The red circle demonstrates the minimizer in this plot.



(a) (b) (c)

0.0.3 Exercise 5.7(a)

I tested four different step lengths, 0.001, 0.01, 0.05 and 0.1. In line with the result, a large step length results in a shorter time to find the minimizer, whereas increasing the possibility of divergence. Plus, there are overshooting and zig-zags in the plot with step length of 0.05 and 0.1. To make a deal between efficiency and convergence, in my tests, a length of 0.01 is the best choice.

0.0.4 Exercise 5.7(e) & (f)

In figure 1(b) and 1(c), the red, green and blue represent the steepest gradient descent, Newton's and BFGS method respectively. From the two figures we can learn that the Newton's method is the first one to converge and it gets closest to the global minimizer. The BFGS's convergence is only worse than that of the Newton's, either in measure of speed or distance to the minimizer. Compared with two other methods, the steepest descent takes a considerable number of iterations to converge, and the result is far from satisfaction.

The steepest descent method has no use of Hession matrix nor its substitutions. Hence it takes the least resources among the three methods, even though the convergence is the worst. It's simple and doesn't need much assumptions to work. The Newton's method is on the other side of the coin: it performs the best in finding minimizer while it's computational expensive. It works under strict assumptions. As a trade-off, the BFGS method is considered to take advantages from the two methods above. Although it is not able to maintain an optimal direction as in the Newton's, it does save a lot of computational resources. In addition, it works under less assumptions than the Newton's.