

Project: Newton and quasi-Newton methods for machine learning

Narindra Razakarivony

February 2022

1 Second order optimization method

You can find the implementation of the method in *src* directory All problem definition and Line search method are in *utils* directory

Question 1.1

Please, refer to the code when dealing with those question.

Starting from two others points the solution is now converging after two iterations

With w_{01} and w_{03} after, respectively 282 iterations and 115 iterations, the function is converging well. However, being far from the origin point make the solution not converging.

The local nature appears in the fact that the starting point should be closed to the solution.

Question 1.3

By applying Global Newton with given parameters and with $\epsilon = 10^{-6}$, the method is converging for w_{01} and w_{03} after **one** iteration but it still struggling with w_{02} .

Question 1.4

For $c = 0.0001$ and $\theta = 0.5$, the time computation is equal to : 0.00010s

For $c = 0.0001$ and $\theta = 1$ the time computation is equal to 0.00013s

For $c = 0.5$ and $\theta = 0.0001$ the time computation is equal to 0.00010s

We notice that lower theta is, time computational become quicker.

Question 1.5

When we apply BGFS on problem 3, the solution is converging 8 iteration from the w^* after one iterations from random points and obviously .

When we apply BGFS on problem 4, the solution is converging 98 iterations with w_{01} and 52 iterations for w_{03} and w_{02}

2 Stochastic second order

Question 2.1/2.2

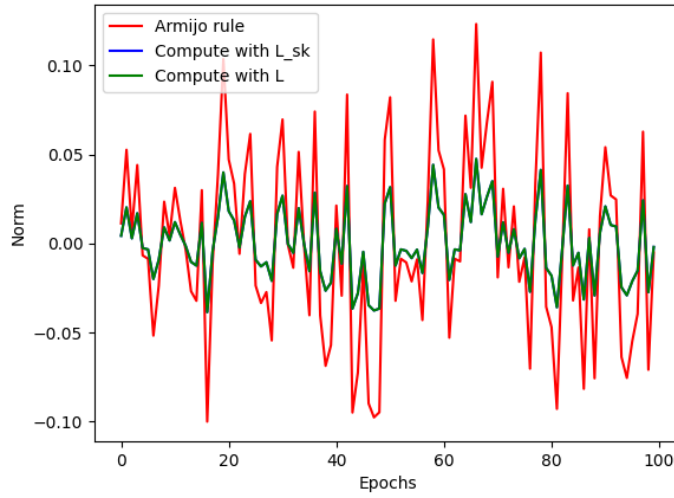


Figure 1: Convergence with different choice of alpha

We have tried to asses the convergence of the sub sampling method with 100 epochs with different step of α .

By choosing $\alpha = \alpha/L_s$ and $\alpha = \alpha/L$, we are converging quickly to the solution. It is less sparse than the Armijo where we it encounters a lot of variations

By choosing $\alpha = 2$, the norm is being greater than the previous one with different choice of α

The BGFS method is not converging with the logit problem for the Armijo Rule. After 100 epochs, the convergence is better with at around 40 - 60 epochs.

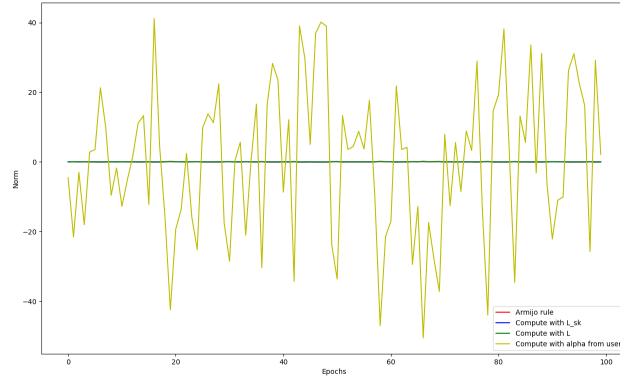


Figure 2: Convergence with different choice of alpha including user choice

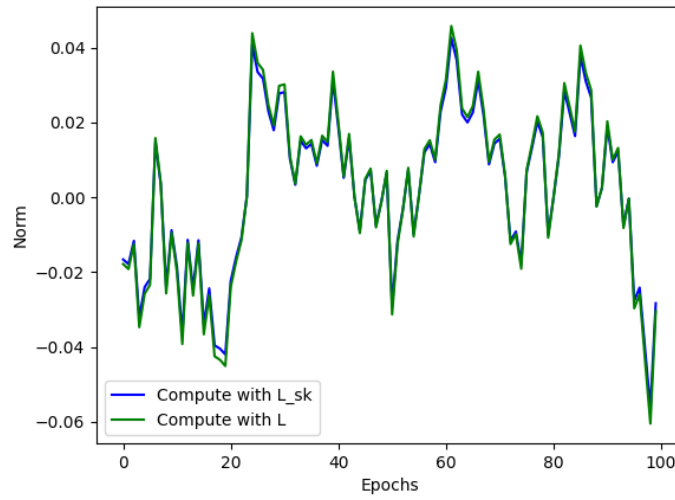


Figure 3: Convergence with Logit problem