The Chinese University of Hong Kong, Shenzhen
SDS · School of Data Science

Fall Semester 2021/22

MDS 6106 – Introduction to Optimization

Exercise Sheet Nr.: 4

Name: Student ID:

In the creation of this solution sheet, I worked together with:

Name: __________ Student ID: __________

Name: __________ Student ID: __________

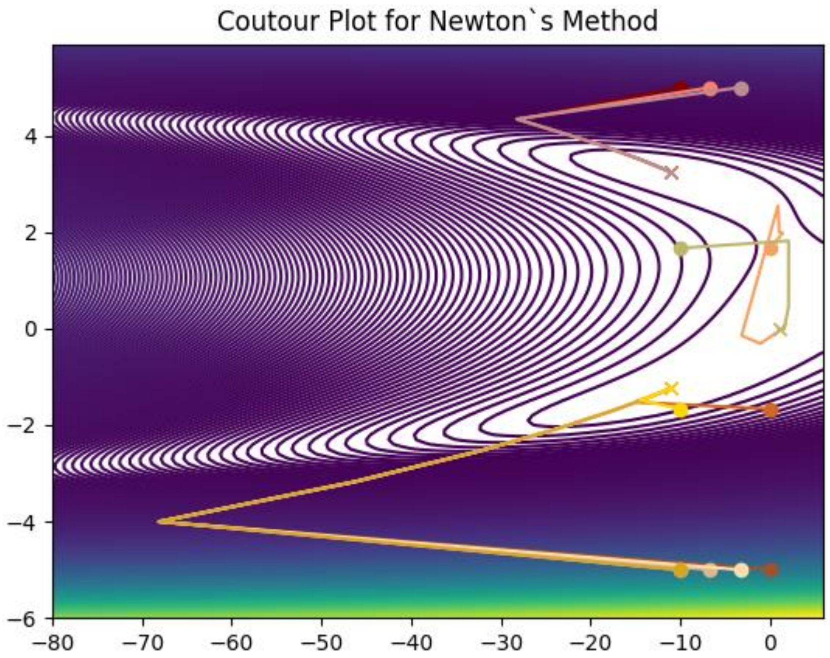
Name: __________ Student ID: __________

For correction:

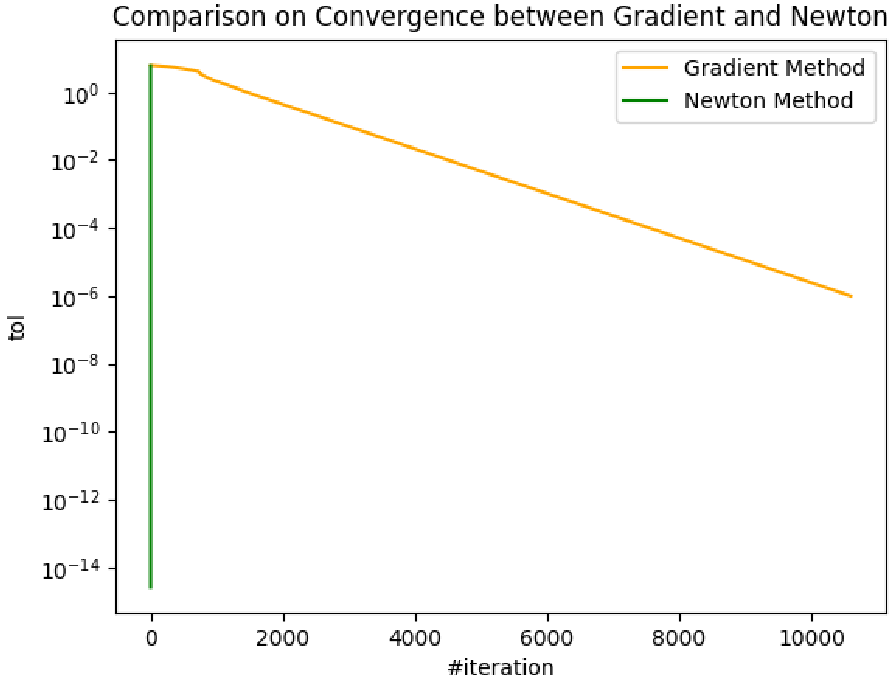
Exercise						$\sum$
Grading						

A1.

a) The solution path of Newton`s method:



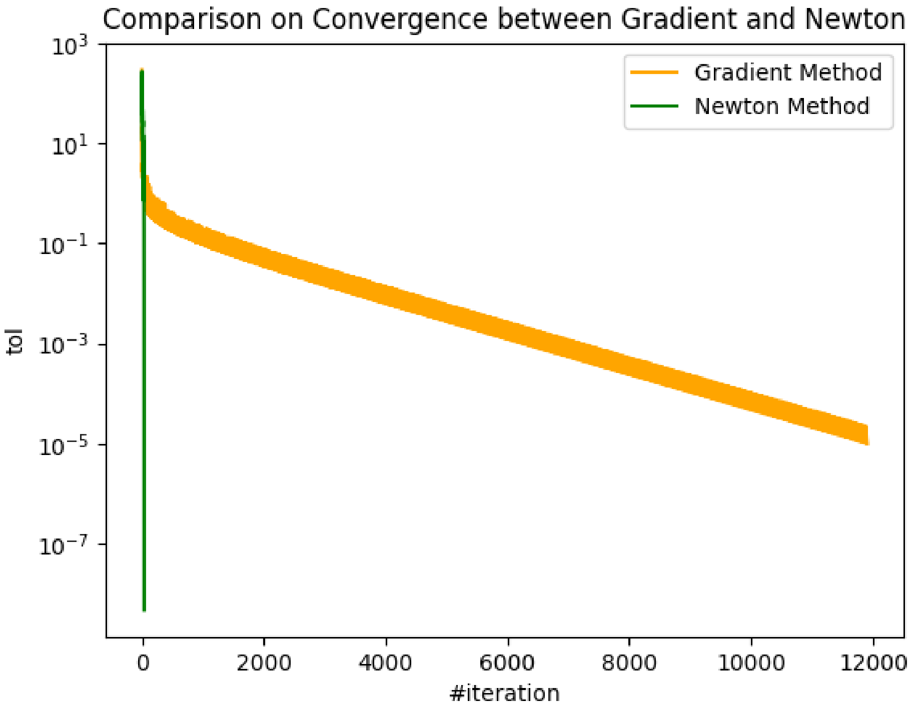
The convergence of 2 methods is visualized in figure below:



It can be seen that Gradient Method follows linear convergence while Newton Method quadrative convergence.

Of the 12 initial points I chose from χ_0, to reach the tolerance of 1e-06, Gradient Method takes 3874.8 rounds to converge, while it only takes 9.7 rounds for Newton Method.

b) The convergence between 2 methods can be demonstrated with figure below:



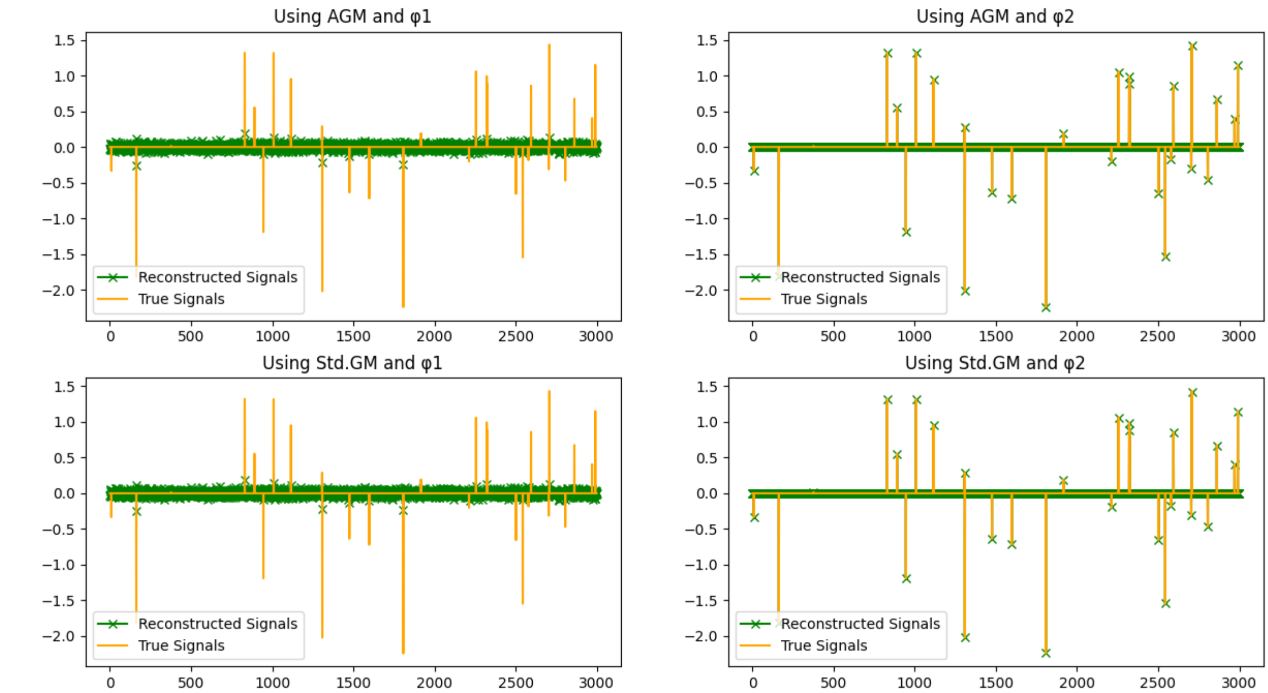
Still Newton Method displays quadratic convergence. However Gradient Method does not exactly follows linear convergence for this function.

From the log generated by the program, the Newton Method always utilize the Newton direction, but not always full step sizes α_k=1. To reach the critical level 0.1 of tolerance, it takes 20 rounds of iteration for Newton Method to converge. During the 20 times of iteration, there 4 times using step sizes smaller than 1. Whereas for Gradient Method, it takes 55 rounds of iteration to converge and it usually takes smaller step sizes, such as 0.001953125. For more information, please refer to the table below:

Tolerance	Rounds of Iteration	
	Newton Method	Gradient Method
1.00E-01	19	54
1.00E-03	20	5231
1.00E-05	21	10916

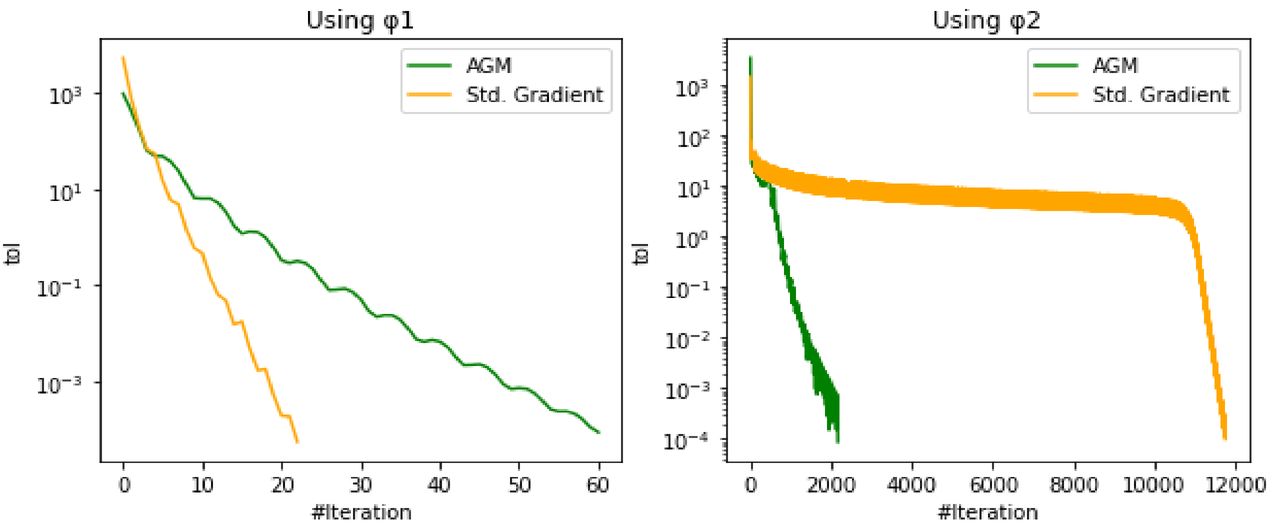
A2.

a) The implementation are finished with Python, whose codes stored in file A42b.py.
b) By measurement of $\|\nabla f(x^k)\|$, the comparison between 2 methods is demonstrated with the following figure:



Whether using Standard Gradient Method or Accelerated Gradient Method, it would be easily seen that Huber norm provides much better recognition of signals.

But as to decide which is best, it takes more information to make the decision.



From the figure above, it could be seen that Accelerated Gradient Method requires less iterations to converge using Huber norm. So the best model would be the combination of Accelerated Gradient Method with Huber norm.

c) Globalized Newton Method should not be applied here.

For $\varphi_1(x)$, the Newton Method should not be applied here since it takes too much memory for the inverse matrix of hessian matrix. For $\varphi_2(x)$, the Huber norm is not twice differentiable to get hessian matrix required by Newton Method.

Instead, the inexact Newton Method, Newton-CG Method or Quasi-Newton Method would be better choices.