Project 2 Optimization

Functions	Function Type	Initial Values	Minimal value	Optimal solution	Achievable Algorithms
Fun 1	Convex	ones(100,1)	0	zeros(100,1)	Newton's, Nelder-mead
Fun 2	Non-Convex	zeros(100, 1)	-2.4432e3	[-1.1094 -1.0697 1.9717 -1.6617 0.4194 …]	Newton's, BFGS
Fun 3	Non-Convex	[3, 3]	2.2775e-16	[1 1]	Newton's (All other methods close but not exact)

Introduction

In this project, we set to implement various common optimization algorithms for several unconstrained optimization problems. We implemented three required algorithms, and two optional algorithms, i.e. conjugate gradient descent and Nelder-Mead. After this project, we have a better understanding of the implementation details of those algorithms.

Functions comparison: Challenges & Solutions

One of the major challenges comes from function 2. This is a rather unwieldy function to define and work with, since all the parameters have to be read from text files. The problem we notice immediately after a test run of optimization algorithms on this function is domain error or complex number result. Upon a closer observation of the function, we notice the logarithm component forces an implicit constraint for the function domain. One solution we propose was that we shrunk the step size until the newly updated input value no longer causes negative input for logarithmic function. Via this approach, we successfully achieve minimal value for the function 2.

Algorithms comparison: Pros & Cons

The gradient descent algorithm was the easiest to implement, and least computationally heavy for one iteration, however, as results showed, it struggled to achieve even a good estimate for two of the three functions.

On the other hand, Newton's method is a very effective algorithm. Due to using Hessian as direction guidance, Newton's method has a much better converge rate compared to gradient descent. Cons are obvious for Newton's method as well. Since it requires a second order derivative, Hessian, the computations are either high or not feasible for not all functions are second order differentiable.

Quasi Newton's methods are alternative to Newton's since they mimic Newton's behavior without requiring a Hessian function. In the project, we implemented BFGS, one of the popular versions of the Quasi Newton's method. It reduces the computational cost and eliminates the needs of Hessian function, however, it no longer can converge as quickly as Newton's method.

Nelder-Mead method is a rather different method compared to other algorithms. The major advantage of this method is that it only requires function and initial value as input. As a gradient free optimization method, the advantage is obvious: it can apply a wide range of functions, ones that are not differentiable. Nonetheless, without the help of gradient or Hessian, Nelder-Mead may take too long to converge.

Conjugate gradient descent is a line search method but for every move, it would not undo part of the moves done previously. It optimizes a quadratic equation in fewer steps than the gradient descent. If x is N-dimensional (N parameters), we can find the optimal point in at most N steps. This is also the fundamental drawback of conjugate gradient descent.

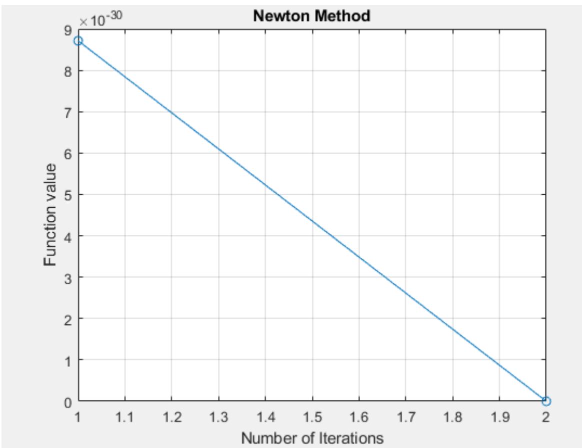
Effects of Initial Values & Parameters

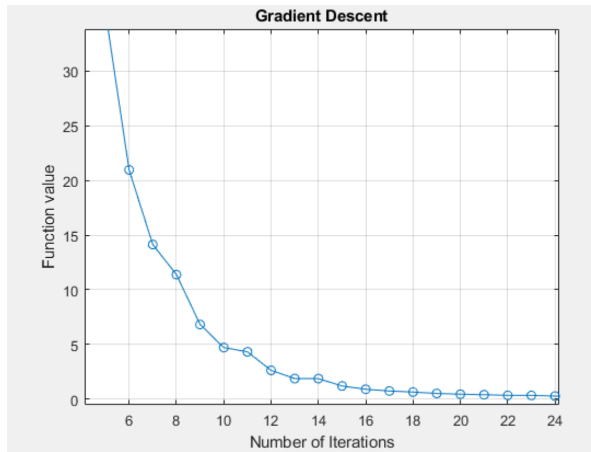
The effects of initial values are quite significant. For an un-convex function, there are multiple local minima, in this case, an unwise choice of initial value may lead to a long time to converge or converge to an nonoptimal local minimum value.

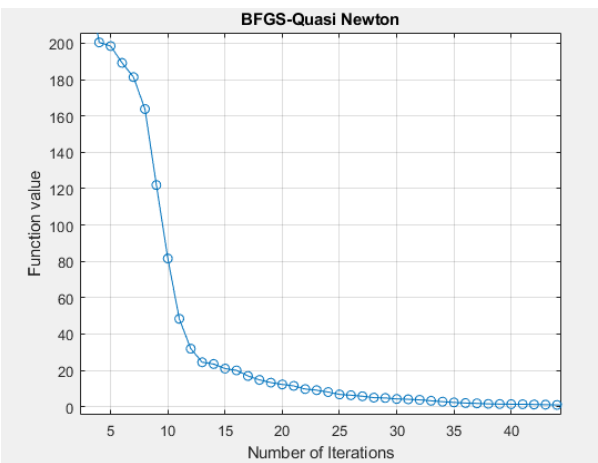
Parameters, such as step size in backtracking linear search are also important. Should the step-size be too large, the algorithm may take much longer time to converge, or having trouble to do so, since the backtracking is heuristic that may not return optimal values.

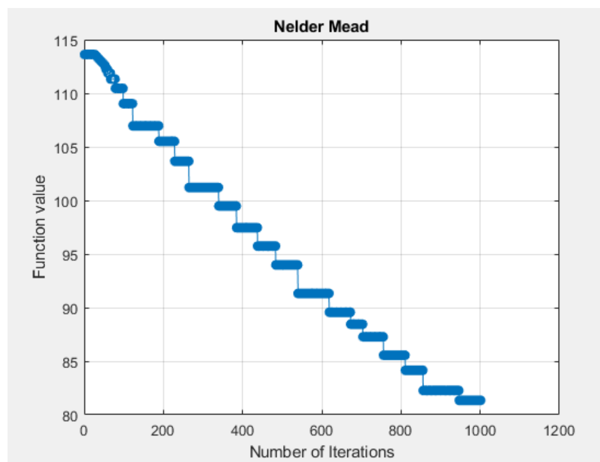
Graphs

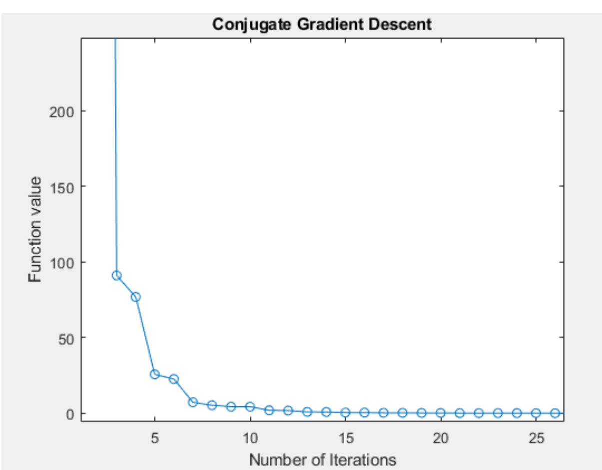
Function 1 -



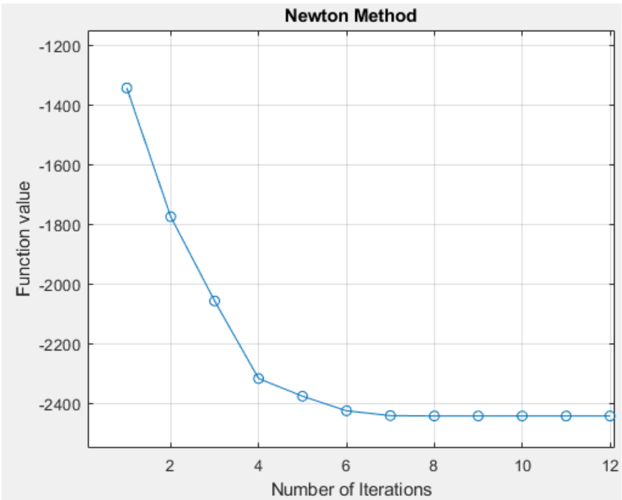




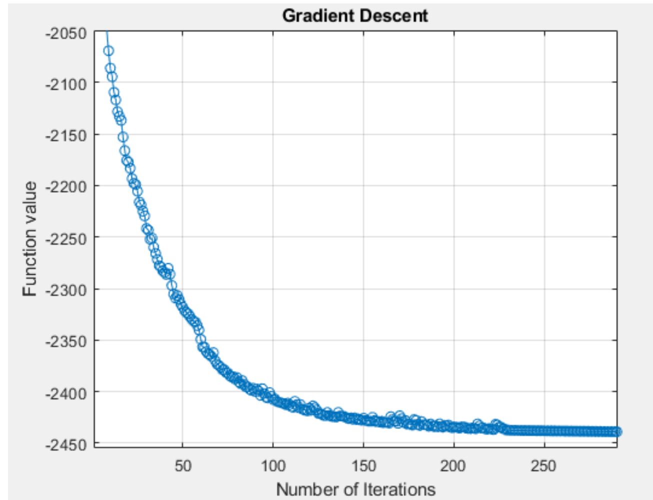




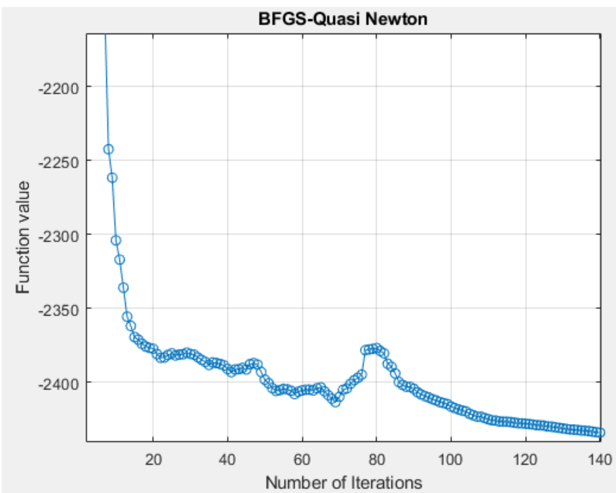
Function 2 -



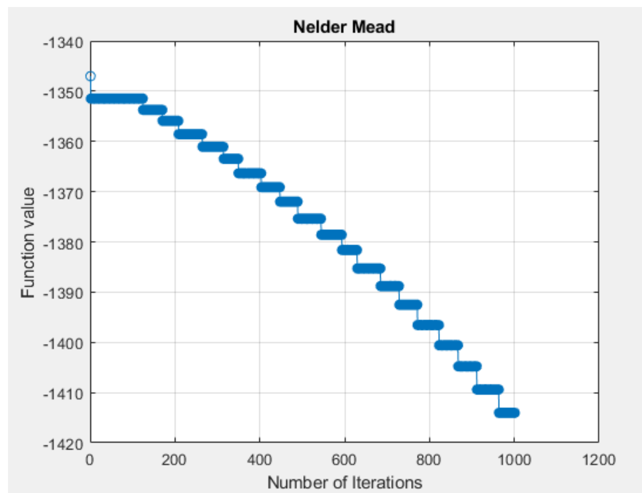
Newton Method
-1200
-1400
-1600
-1800
-2000
-2200
-2400
Function value
2
4
6
8
10
12
Number of Iterations



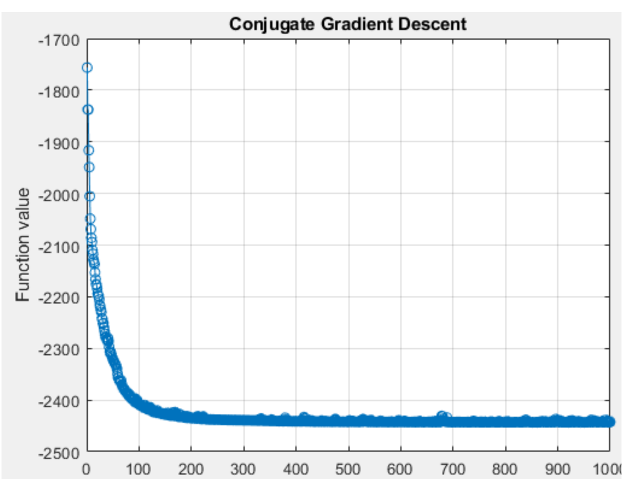
Gradient Descent
-2050
-2100
-2150
-2200
-2250
-2300
-2350
-2400
-2450
Function value
50
100
150
200
250
Number of Iterations



BFGS-Quasi Newton
-2200
-2250
-2300
-2350
-2400
Function value
20
40
60
80
100
120
140
Number of Iterations

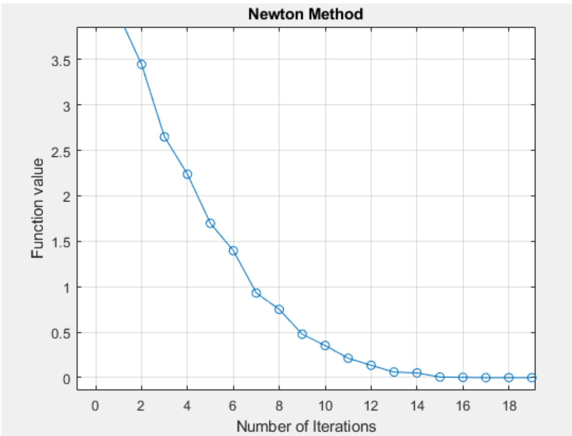


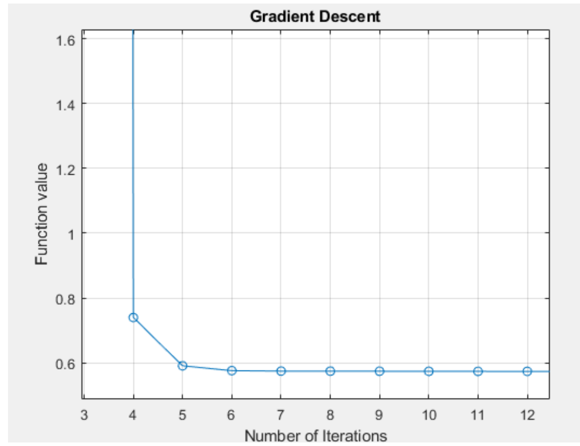
Nelder Mead
-1340
-1350
-1360
-1370
-1380
-1390
-1400
-1410
-1420
Function value
0
200
400
600
800
1000
1200
Number of Iterations

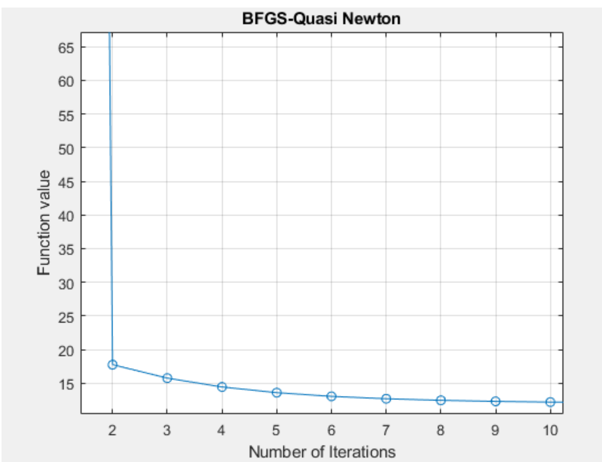


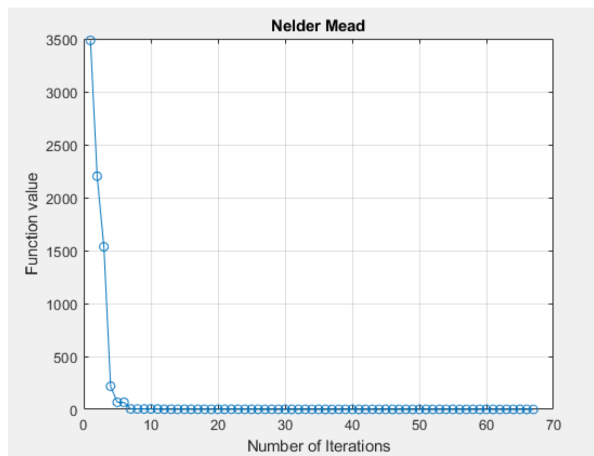
Conjugate Gradient Descent
-1700
-1800
-1900
-2000
-2100
-2200
-2300
-2400
-2500
Function value
0
100
200
300
400
500
600
700
800
900

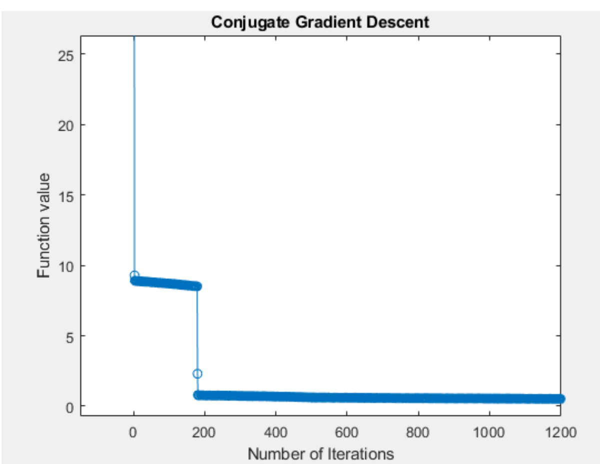
Function 3 -


Newton Method
Function value
Number of Iterations


Gradient Descent
Function value
Number of Iterations


BFGS-Quasi Newton
Function value
Number of Iterations


Nelder Mead
Function value
Number of Iterations


Conjugate Gradient Descent
Function value
Number of Iterations