**EECS 498 HW2**

**Questions**

**Prob. 1**

Define a point f in n-dimension space:

**Prob. 2**

is convex since it is a combination of hyperplane

is convex since it is a hyperplane

is convex since

is convex since it is a hyperplane

Since are convex and operation of sum of convex functions preserves convexity

is convex

**Prob. 3**

By Symmetric Difference Quotient

Numerical results with are smaller than the analytical result, except the result with and it is also the closest result to the analytical one.

**Prob. 4**

**Prob. 5**

There would be more than one subgradient when intersects

**Prob. 6**

1. Lagrange dual function

Lagrange function:

Dual function:

is a vector of Lagrange multipliers

1. Dual problem

Maximize

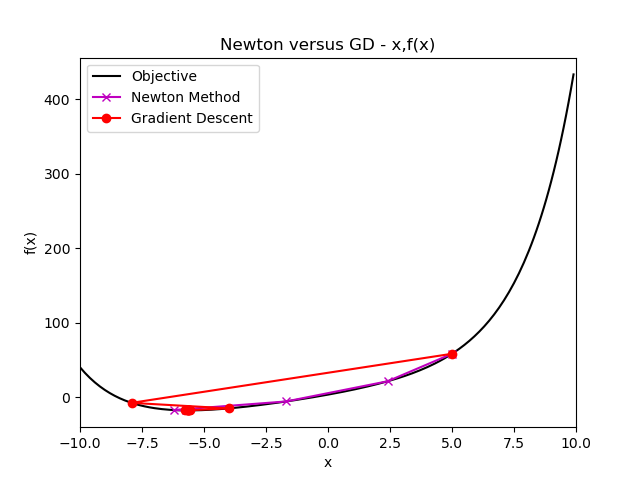
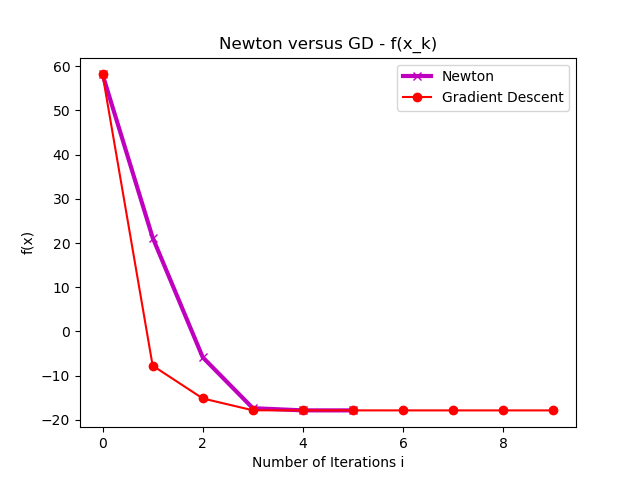
Subject to

1. Optimal value of dual problem relation with primal optimal?

Since we assume primal is feasible and bounded, then the strong duality holds.

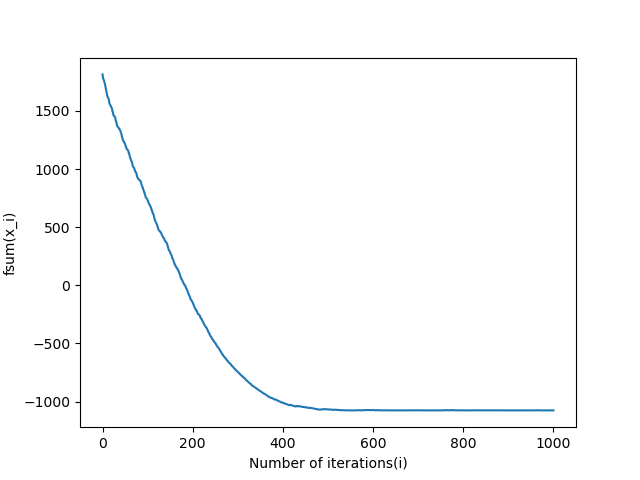
**Implementation**

**Prob 1d**

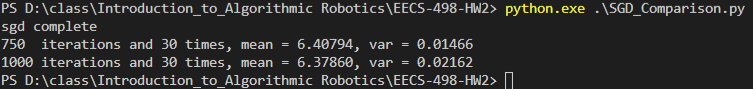
In terms of total iterations, Newton method is better than Gradient Descent in this case. Since convergence rate of Gradient Descent is greatly depending on the condition number of Hessian, and we can observe that the condition number of the hessian around optimal is high(not sensitive), Gradient Descent would take more iterations to reach termination condition. In comparison, Newton method doesn’t have this problem by utilizing second order approximation.

**Prob 2b**



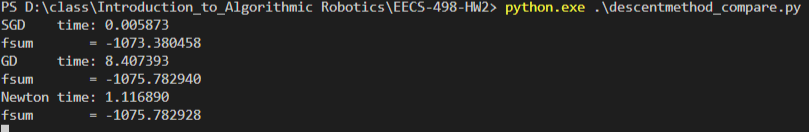
is not always descending. SGD takes a random term of derivative fi to update x every iteration, which means it doesn’t guarantee that the direction of is always toward the descending direction.

**Prob 2c**



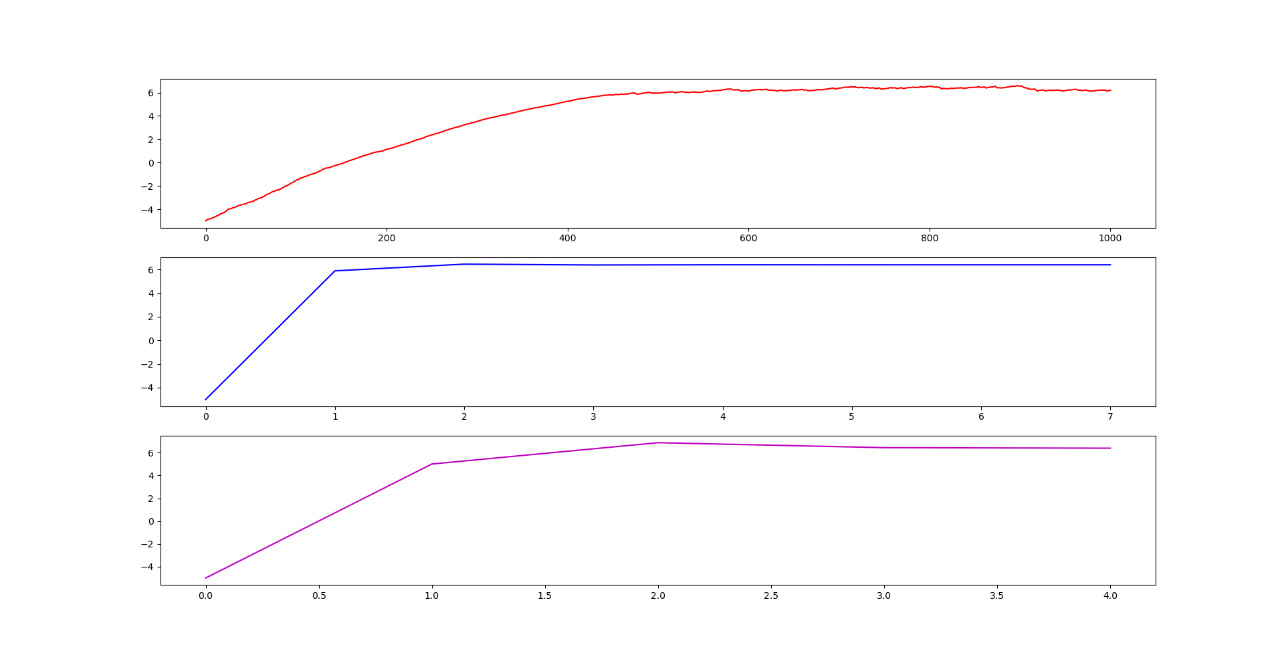
The result from 1000 iterations is more various and far away from optimal than the result of 750 iterations. The phenomenon results from the stochastic sense of SGD, which doesn’t guarantee that the more iterations, the closer to the optimal and less various.

**Prob 2d**



1. SGD only computes one derivative of fi each iteration, which means the computation cost of SGD with the setting is proportional to the number of iterations (1000).

GD and Newton method, on the other hand, need to compute the derivative of current x (and even Hessian of x in terms of Newton method). In this case, there are 10000 of fi.

  
From Prob 1, we know that GD may suffer from the condition number of f and takes more iterations. In the experiment, GD takes 7 iterations and Newton method takes 4.

1. In terms of , performance of GD and Newton method are about the same, and SGD is the worst. As mention in Prob. 2b, since SGD randomly pick a term to update x, direction of doesn’t always toward the descent direction by SGD. Update by SGD may bouncing around the optimal.

**Prob 3a**

: scale for objective force  
 : objective function from the primal problem  
: inequality constraints

**Prob 3b**

: coefficients of objective functions  
: coefficients of inequality constraints  
: constants values of inequality constraints

**Prob 3c**

**Prob 3d**

**Prob 3e**

**Prob 4a**

**Prob 4b**

Minimize

Subject to