Q1. Exercise 8

8. For $\delta > 0$ define the map $\mathbf{S}^\delta$ by

$$\mathbf{S}^\delta(\mathbf{x},\ \mathbf{d}) = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \alpha\mathbf{d}, \quad 0 \le \alpha \le \delta; \quad f(\mathbf{y}) = \min_{0 \le \beta \le \delta} f(\mathbf{x} + \beta\mathbf{d})\}.$$

Thus $\mathbf{S}^\delta$ searches the interval $[0,\ \delta]$ for a minimum of $f(\mathbf{x} + \alpha\mathbf{d})$, representing a "limited range" line search. Show that if f is continuous, $\mathbf{S}^\delta$ is closed at all $(\mathbf{x},\ \mathbf{d})$.

Solution:

Assume $\mathbf{x_k} \to \mathbf{x}^*$, $\mathbf{d_k} \to \mathbf{d}^*$, $\mathbf{y_k} = \mathbf{x_k} + \alpha_k \mathbf{d_k} \in \mathbf{S}^\delta(\mathbf{x_k},\ \mathbf{d_k})$, $\mathbf{y}_k \to \mathbf{y}^*$.
If $\mathbf{d}^* \neq \mathbf{0}$, we have

$$\alpha_k = \frac{\|\mathbf{y}_k - \mathbf{x}_k\|}{\|\mathbf{d}_k\|} \to \frac{\|\mathbf{y}^* - \mathbf{x}^*\|}{\|\mathbf{d}^*\|} \triangleq \alpha^*.$$

Since $\alpha_k \in [0, \delta]$, we have $\alpha^* \in [0, \delta]$.
If $\mathbf{d}^* = \mathbf{0}$. Since $\alpha_k \in [0, \delta]$, $\mathbf{d}_k \to \mathbf{0}$, we have

$$\mathbf{y}_k = \mathbf{x_k} + \alpha_k \mathbf{d_k} \to \mathbf{x}^*.$$

So we have $\mathbf{y}^* = \mathbf{x}^* + \alpha^* \mathbf{d}^*$ for some $\alpha^* \in [0, \delta]$ in either case.
For any k, $\beta \in [0, \delta]$, we have

$$f(\mathbf{y}_k) \le f(\mathbf{x}_k + \beta \mathbf{d}_k)$$

by the definition of $\mathbf{y}_k$. Since f is continuous, taking the limit of both sides, we have

$$f(\mathbf{y}^*) \le f(\mathbf{x}^* + \beta \mathbf{d}^*)$$

for any $\beta \in [0, \delta]$. So

$$f(\mathbf{y}^*) \le \min_{0 \le \beta \le \delta} f(\mathbf{x}^* + \beta \mathbf{d}^*),$$

which concludes $\mathbf{y}^* \in \S^\delta(\mathbf{x}^*,\ \mathbf{d}^*)$. So $\S^\delta$ is closed at all $(\mathbf{x}, \mathbf{d})$.

Q2. Exercise 9

Assume $\mathbf{x_k} \to \mathbf{x}^*$, $\mathbf{d_k} \to \mathbf{d}^*$, $\mathbf{y_k} = \mathbf{x_k} + \alpha_k \mathbf{d_k} \in {}^{\varepsilon}\S(\mathbf{x_k},\ \mathbf{d_k})$, $\mathbf{y}_k \to \mathbf{y}^*$.
Since $\mathbf{d}^* \neq \mathbf{0}$, we can assume $\mathbf{d}_k \neq \mathbf{0}$ for all sufficiently large k. Therefore,

$$\alpha_k = \frac{\|\mathbf{y}_k - \mathbf{x}_k\|}{\|\mathbf{d}_k\|} \to \frac{\|\mathbf{y}^* - \mathbf{x}^*\|}{\|\mathbf{d}^*\|} \triangleq \alpha^*.$$

So $\mathbf{y}^* = \mathbf{x}^* + \alpha^* \mathbf{d}^*$.
For any k and $\beta \geq 0$, we have

$$f(\mathbf{y}_k) \leq f(\mathbf{x}_k + \beta \mathbf{d}_k) + \varepsilon$$

Taking the limit, we have

$$f(\mathbf{y}^*) \leq f(\mathbf{x}^* + \beta \mathbf{d}^*) + \varepsilon, \forall \beta \geq 0.$$

So, we have

$$f(\mathbf{y}^*) \leq \min_{\beta \geq 0} f(\mathbf{x}^* + \beta \mathbf{d}^*) + \varepsilon$$

Therefore, $\mathbf{y}^* \in {}^{\varepsilon}\S(\mathbf{x}^*, \mathbf{d}^*)$, which means ${}^{\varepsilon}\S$ is closed at all $(\mathbf{x}, \mathbf{d})$ with $\mathbf{d} \neq 0$.

Q3. For MATLAB codes, please see the HW4_solution folder at Canvas. I strongly encourage you to try them if you had difficult time implementing the algorithms. The codes have everything you need to regenerate the results in this solution.

(a) ***<u>Gradient Descent with Armijo's rule for inexact line search.</u>*** In the following, we have (i) the plot of the Euclidean norm of gradient versus the iteration index, (ii) the plot of objective function value versus iteration index, and (iii) the plot of the Euclidean distance from the solution point to the local minimum point (i.e., (x, y) = (20, 3)) versus index. In the third plot, you can see that the solution point converges to the local minimum point linearly; you can see that

$$\log(||x_k - x_{opt}||) \sim \beta_0 - \beta k$$

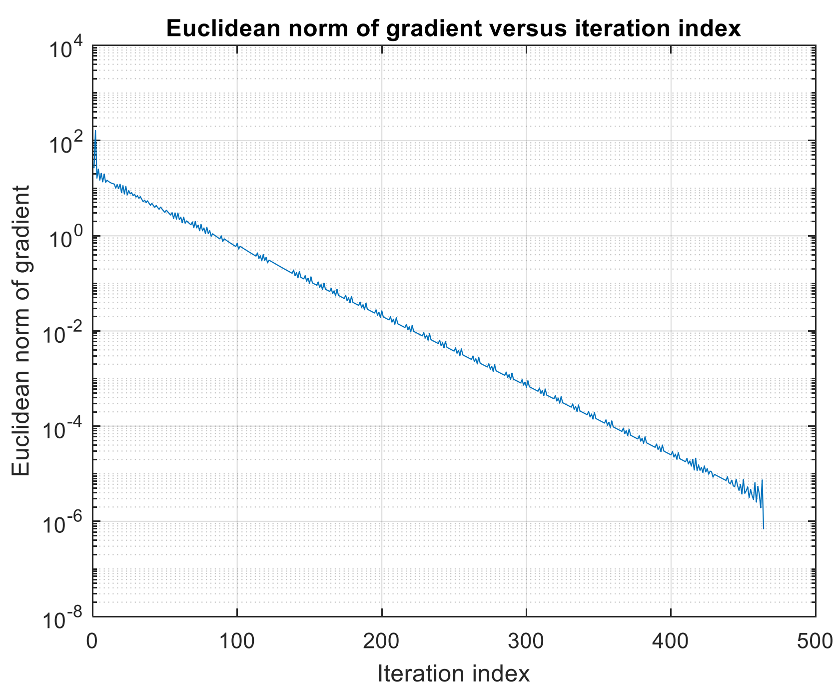
with a positive β. This implies that

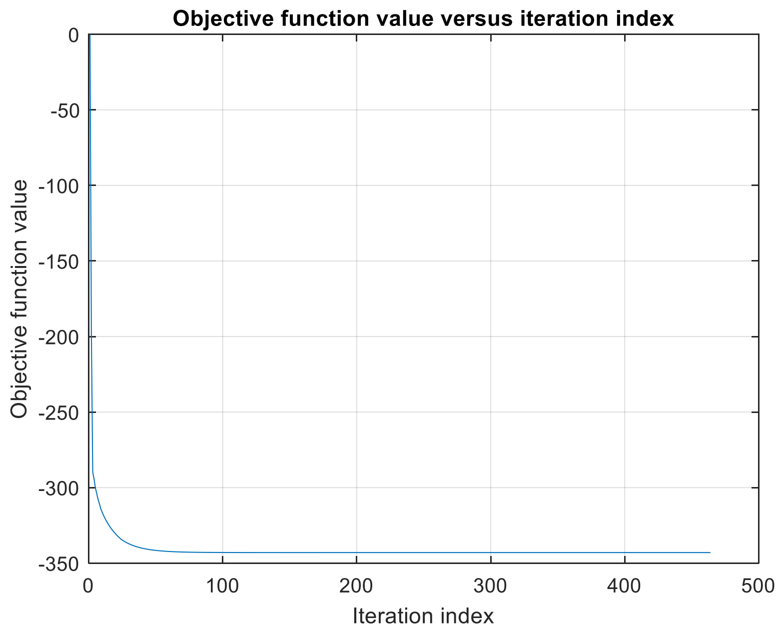
$$||x_k - x_{opt}|| \sim C\left(e^{-\beta}\right)^k$$

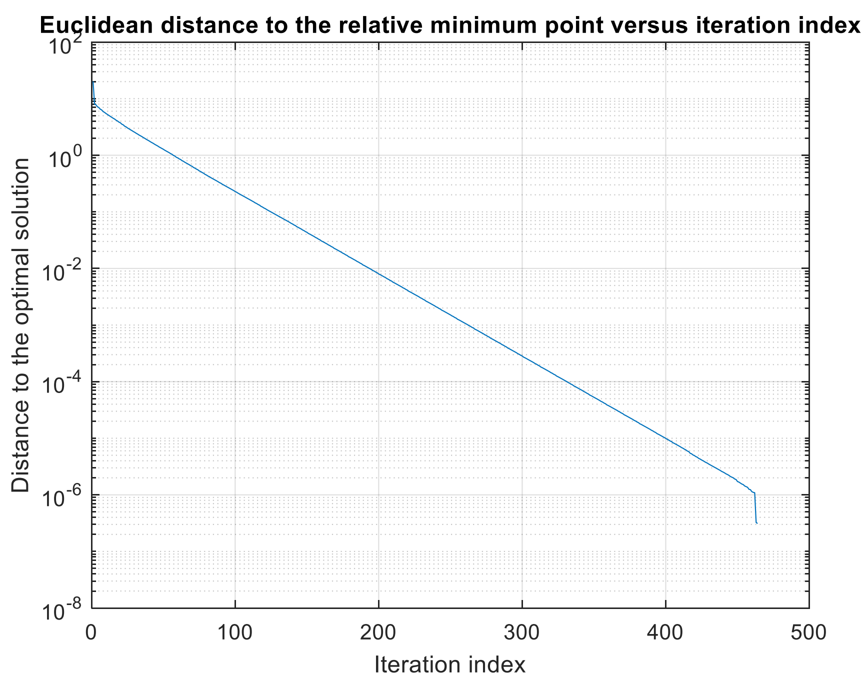
with $e^{-\beta}$ between 0 and 1. Therefore, the plot of the error function value suggests the linear convergence of Gradient Descent with Armijo's rule. The Hessian at the final solution x_{final} was found as

$$\begin{bmatrix} 2 & -5 \\ -5 & 108 \end{bmatrix}$$

which is a **positive definite** matrix. And, the gradient has almost zero magnitude (according to the second plot below) at x_{final}. Therefore, we can conclude that x_{final} is a local relative minimum.

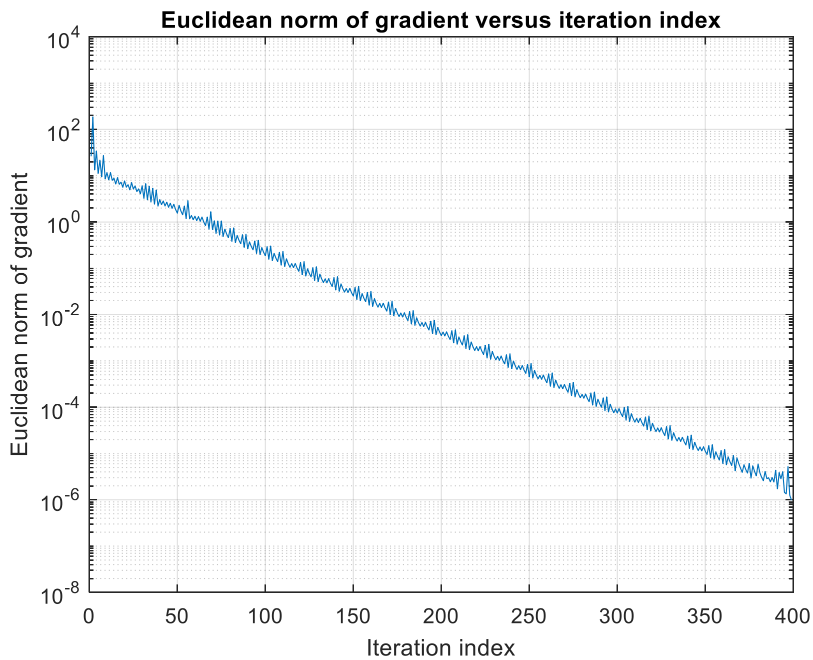


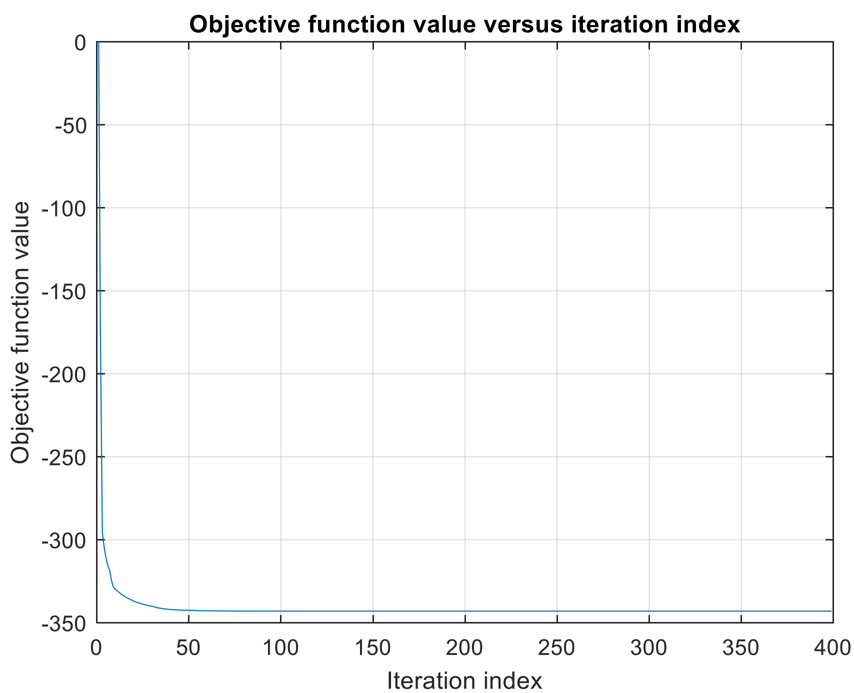


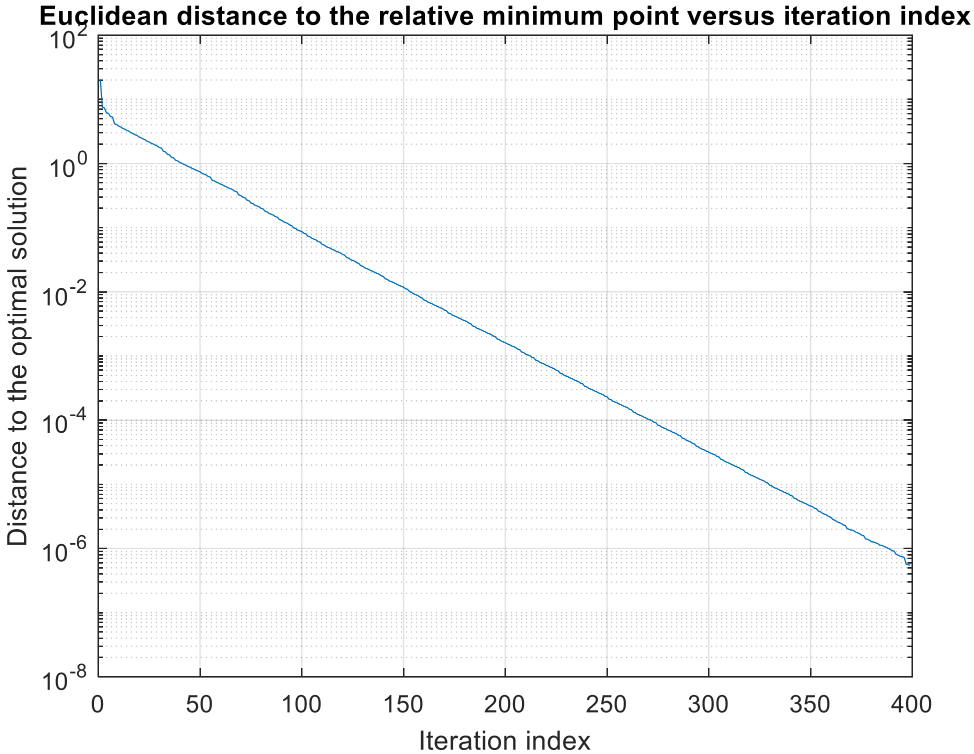


(b) ***Gradient Descent with Goldstein rule for inexact line search.*** Similar to part (a), in the following, we have (i) the plot of the Euclidean norm of gradient versus the iteration index, (ii) the plot of objective function value versus iteration index, and (iii) the plot of the Euclidean distance from the solution point to the local minimum point (i.e., (x, y) = (20, 3)) versus index.

You can observe almost same convergence behaviors as in the figures of part (a). From the third plot, we can observe linear local convergence of this algorithm. The Hessian the final solution x_{final} was found to be the same positive definite matrix as in part (a). Therefore, we can be assured that x_{final} is a relative minimum point.


Euclidean norm of gradient versus iteration index
Euclidean norm of gradient
Iteration index
10^4
10^2
10^0
10^-2
10^-4
10^-6
10^-8
0
50
100
150
200
250
300
350
400


Objective function value versus iteration index
Objective function value
Iteration index
0
-50
-100
-150
-200
-250
-300
-350
0
50
100
150
200
250
300
350
400


Euclidean distance to the relative minimum point versus iteration index
Distance to the optimal solution
Iteration index
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
0
50
100
150
200
250
300
350
400