1. The BFGS method iterates as \( x_{k+1} = x_k - \alpha_k H_k^{-1} \nabla f(x_k) \). The matrix \( H_k \) is iterated by the formula
\[
H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k}
\]
where \( s_k = x_{k+1} - x_k \) and \( y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \). When implementing the BFGS method, it will be better to directly update \( H_k^{-1} \) other than first obtaining \( H_k \) and then solving \( H_k^{-1} \nabla f(x_k) \). Your task is to show the following iteration formula for \( H_k^{-1} \) by manipulating (1):
\[
H_{k+1}^{-1} = \left( I - \frac{s_k y_k^T}{y_k^T s_k} \right) H_k^{-1} \left( I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \frac{s_k s_k^T}{y_k^T s_k}
\]
(Hint: Use the Sherman-Morrison-Woodbury formula (or the so-called matrix inversion lemma): \((A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + VT A^{-1}U)^{-1}VT A^{-1}\) where \( A \in \mathbb{R}^{n \times n}, U,V \in \mathbb{R}^{n \times d}\) are matrices such that \( A + UV^T \) is nonsingular.)

2. Programming Assignment
(a) First, you are asked to implement the gradient method with Armijo rule and the BFGS method to solve the positive definite quadratic minimization problem:
\[
\min_{x \in \mathbb{R}^p} \frac{1}{2} x^T Q x + q^T x + r
\]
where \( Q \) is positive definite. Generate \( Q, q, \) and \( r \) via the same code that you used to generate \( Q, q, \) and \( r \) in Homework 2. For the gradient method, you should experiment the Armijo rule and compare it with the constant stepsize case where \( \alpha = \frac{1}{L} \). Fix \( \alpha_0 = \frac{2}{L} \). Try a few choices of \( \sigma \) and \( \beta \) and plot the best case. For the BFGS method, also use the Armijo rule. Specifically, choose \( \alpha = \alpha_0 \beta^m \) where \( m \) is the smallest integer such that
\[
f(x_k - \alpha_0 \beta^m H_k^{-1} \nabla f(x_k)) \leq f(x_k) - \sigma \alpha_0 \beta^m \nabla f(x_k)^T H_k^{-1} \nabla f(x_k)
\]
Fix \( \alpha_0 = 1 \). Try a few choices of \( \beta \) and \( \sigma \) and plot the best case. Always start from the initial condition \( x_0 = x_{-1} = (1; 1; \ldots; 1)^T \). You are asked to turn in plots of the progression of objective values (relative to the minimum) for \( p = 200 \) and various choices of \((m, L)\) values \((m = 1, L = 10; m = 0.1, L = 1000)\). Notice for this quadratic problem, the optimal point
$x^* = -Q^{-1}q$ can be directly computed when the dimension $p$ is not that high. This can be used when you plot the progression of objective values relative to the minimum. The $y$ axis for your plots should be in log scale. You can try different iteration number (e.g. $k = 1000$) until the algorithm converges. Then briefly discuss your findings.

(b) Here you are asked to implement the gradient method and the BFGS method for a non-convex function. Consider the Rosenbrock function

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$$

This function has a unique stationary point at $(1, 1)$ which is also the global min. Implement the gradient method and the BFGS method with Armijo rule. Now you are asked to experiment various choices of $\alpha_0$, $\sigma$, and $\beta$. First start with the initial condition $(1.2, 1.2)$ and then try another initial condition $(-1.2, 1)$. Does the initial condition matter? Which method performs better? Discuss your findings.