1. Consider the following algorithm for minimizing a function f:

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)},$$

where

$$\alpha_k = \arg\min_{\alpha} f\big(x^{(k)} + \alpha d^{(k)}\big).$$

Let $g^{(k)} = \nabla f\big(x^{(k)}\big)$ (as usual).
Suppose that f is quadratic with Hessian Q. We choose

$$d^{(k+1)} = \gamma_k g^{(k+1)} + d^{(k)},$$

and we wish the directions $d^{(k)}$ and $d^{(k+1)}$ to be Q-conjugate. Find a formula for γ_k in terms of $d^{(k)}$, $g^{(k+1)}$, and Q.

Solution:
In order to have the directions $d^{(k)}$ and $d^{(k+1)}$ Q-conjugate, we must have

$$(d^{(k)})^T Q d^{(k+1)} = 0$$

$$(d^{(k)})^T Q(\gamma_k g^{(k+1)} + d^{(k)}) = 0$$
$$(d^{(k)})^T Q \gamma_k g^{(k+1)} + (d^{(k)})^T Q d^{(k)} = 0$$
$$\gamma_k = \frac{-(d^{(k)})^T Q d^{(k)}}{(d^{(k)})^T Q g^{(k+1)}}$$

2. Represent the function

$$f(x) = \frac{5}{2}x_1^2 + x_2^2 - 3x_1x_2 - x_2 - 7$$

in the form

$$f(x) = \frac{1}{2}x^\top Qx - x^\top b + c.$$

Then, use the conjugate gradient algorithm to find a minimizer with $d^{(0)} = \nabla f(x^{(0)})$, where $x^{(0)} = 0$.

Solution:
The gradient of this function is $\nabla f(x) = Qx - b$.

$$\nabla f(x) = \begin{bmatrix} 5x_1 - 3x_2 \\ 2x_2 - 3x_1 - 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} x - x^\top [0, 1]^\top - 7$$

The conjugate gradient algorithm is

$$\begin{aligned}
d^{(0)} &= \nabla f(x^{(0)}) \\
g^{(0)} &= \nabla f(x^{(0)}) \\
\alpha_k &= -\frac{(g^{(k)})^T d^{(k)}}{(d^{(k)})^T Q d^{(k)}} \\
x^{(k+1)} &= x^{(k)} - \alpha_k d^{(k)} \\
g^{(k+1)} &= \nabla f(x^{(k+1)}) \\
\beta_k &= \frac{g^{(k+1)})^T Q d^{(k)}}{(d^{(k)})^T Q d^{(k)}} \\
d^{(k+1)} &= -g^{(k+1)} + \beta_k d^{(k)}
\end{aligned}$$

In practice, we know we have found the minimizer if $g^{(k+1)} = 0$. To complete this problem, simply run the above algorithm with $d^{(0)} = \nabla f(x^{(0)}) = [0, -1]^\top$ and $x^{(0)} = [0, 0]^\top$.

$$\begin{aligned}
\alpha_0 &= -\frac{\begin{bmatrix} 0 \\ -1 \end{bmatrix}^T \begin{bmatrix} 0 \\ -1 \end{bmatrix}}{\begin{bmatrix} 0 \\ -1 \end{bmatrix}^T \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix}} = -\frac{1}{2} \\
x^{(1)} &= x^{(0)} - \frac{1}{2} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} \\
g^{(1)} &= \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} - \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ 0 \end{bmatrix} \\
\beta_0 &= -\frac{9}{4}
\end{aligned}$$

3. Consider the DFP algorithm applied to the quadratic function

$$f(x) = \frac{1}{2} x^\top Q x - x^\top b,$$

where $Q = Q^\top > 0$.

(a) Write down a formula for α_k in terms of Q, $g^{(k)}$, and $d^{(k)}$.

(b) Show that if $g^{(k)} \neq 0$, then $\alpha_k > 0$.

Solution:
By the DFP algorithm, $\alpha_k = argmin f(x^{(k)} + \alpha_k d^{(k)})$

$$\begin{aligned}
\phi(\alpha) &= \frac{1}{2}(x^{(k)} + \alpha d^{(k)})^\top Q (x^{(k)} + \alpha d^{(k)}) - (x^{(k)} + \alpha d^{(k)})^\top b \\
\phi'(\alpha) &= (d^{(k)})^\top Q x^{(k)} + \alpha (d^{(k)})^\top Q d^{(k)} - b^\top d^{(k)} = 0 \\
\alpha &= \frac{-d^{(k)}(Qx^{(k)} - b)}{(d^{(k)})^\top d^{(k)}} = \frac{(-d^{(k)})^\top g^{(k)}}{(d^{(k)})^\top d^{(k)}}
\end{aligned}$$

Recall that by DFP, we have

$$d^{(k)} = -H_k g^{(k)}$$

So, we have

$$\alpha = \frac{(g^{(k)})^\top H_k g^{(k)}}{(d^{(k)})^\top d^{(k)}}$$

With the above equation, it is simple to see that $\alpha > 0$ if $g^{(k)} > 0$ and if $g^{(k)} < 0$.
So we can conclude that if $g^{(k)} \neq 0$, then $\alpha > 0$.

4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $f \in C^1$. Consider an optimization algorithm applied to this f, of the usual form

$$x^{(k+1)} = x^{(0)} + \alpha_k d^{(k)},$$

where $\alpha_k \geq 0$ is chosen according to line search. Suppose that

$$d^{(k)} = -H_k g^{(k)},$$

where $g^{(k)} = \nabla f(x^{(k)})$ and H_k is symmetric.

(a) Show that if H_k satisfies the following conditions whenever the algorithm is applied to a quadratic, then the algorithm is quasi-Newton:

i. $H_{k+1} = H_k + U_k$.
ii. $U_k \Delta g^{(k)} = \Delta x^{(k)} - H_k \Delta g^{(k)}$.
iii. $U_k = a^{(k)}(\Delta x^{(k)})^\top + b^{(k)}(\Delta g^{(k)})^\top H_k$, where $a^{(k)}$ and $b^{(k)}$ are in $\mathbb{R}^n$.

(b) Which (if any) among the rank-one, DFP, and BFGS algorithms satisfy the three conditions in part (a) (whenever the algorithm is applied to a quadratic)? For those that do, specify the vectors $a^{(k)}$ and $b^{(k)}$.