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

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

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

2. Represent the function

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

in the form

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

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

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

The stopping condition of the conjugate gradient algorithm is $g^{(k+1)} = 0$.
To complete this problem, simply run the above algorithm with $d^{(0)} = \nabla f(x^{(0)}) = [0, -1]^\top$ and $x^{(0)} = [0, 0]^\top$.

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

Continue until $g^{(k+1)} = 0$

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 $b^{(k)}$.

(a) Quasi Newton algorithms have the form

$$d^{(k+1)} = -H_k g^{(k+1)} \tag{1}$$

$$\alpha_k = argmin_{alpha \geq 0} f(x^k + \alpha d^k) \tag{2}$$

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \tag{3}$$

In the quadratic case, the above matrices are required to satisfy

$$H_{k+1} \Delta g^{(i)} = \Delta x^{(i)}, 1 \leq i \leq k \tag{4}$$

We want to show (4) holds
Using the given equations, we have

$$U_k = H_{k+1} - H_k \tag{5}$$

$$U_k \Delta g^{(k)} = (H_{k+1} - H_k)\Delta g^{(k)} = \Delta x^{(k)} - H_k \Delta g^{(k)} \tag{6}$$

$$H_{k+1} \Delta g^{(k)} = \Delta x^{(k)} \tag{7}$$

To show that this holds for all $1 \leq i \leq k$, we use induction: Assume that the above result holds for $k-1$, then

$$\begin{aligned}
H_{k+1}\Delta g^{(k-1)} &= H_{k-1}\Delta g^{(k-1)} + U_{k-1}\Delta g^{(k-1)} && (8)\\
&= \Delta x^{(k-1)} + U_{k-1}\Delta g^{(k-1)}(IH) && (9)\\
&= \Delta x^{(k-1)} + (a^{(k)}(\Delta x^{(k)})^\top + b^{(k)}(\Delta g^{(k)})^\top H_k)\Delta g^{(k-1)} && (10)\\
&= \Delta x^{(k-1)} + a^{(k)}(\Delta x^{(k)})^\top \Delta g^{(k-1)} + b^{(k)}(\Delta g^{(k)})^\top H_k \Delta g^{(k-1)} && (11)
\end{aligned}$$

Now we want to show that the right two terms of (11) are equal to 0.
To do this, we will use Theorem 11.1 which states (simplified) that for a quasi-Newton algorithm applied to a quadratic function, the directions $d^{(0)}, ..., d^{(k+1)}$ are Q-conjugate Using this theorem, and condition (3), we can find that

$$\begin{aligned}
a^{(k)}(\Delta x^{(k)})^\top \Delta g^{(k-1)} &= a^{(k)}(\alpha_k d^{(k)})^\top Q \Delta x^{(k-1)}\\
&= a^{(k)}\alpha_k (d^{(k)})^\top Q d^{(k-1)}\alpha_{k-1}\\
&= 0
\end{aligned}$$

Repeat a similar simplification for the b term to find that it is equal to 0.
Therefore, by induction, we can conclude that for all $1 \leq i \leq k$, $H_{k+1}\Delta g^{(i)} = \Delta x^{(i)}$, which completes the proof.

(b) To show that the first two conditions hold, we want to show that

$$H_{k+1}\Delta g^{(i)} = \Delta x^{(i)}.$$

Rank-one Algorithm:
By Theorem 11.2, for the rank-one algorithm applied to the quadratic with Hessian $Q = Q^\top$, we have

$$H_{k+1}\Delta g^{(k)} = \Delta x^{(k)}, \quad 0 \leq k < l.$$

Thus, the first two conditions are satisfied.
For condition 3:

$$U_k = H_{k+1} - H_k = \frac{(\Delta x^{(k)} - H_k\Delta g^{(k)})(\Delta x^{(k)} - H_k\Delta g^{(k)})^\top}{\Delta g^{(k)\top}(\Delta x^{(k)} - H_k\Delta g^{(k)})}.$$

From this:

$$a^{(k)} = \frac{\Delta x^{(k)} - H_k\Delta g^{(k)}}{\Delta g^{(k)\top}(\Delta x^{(k)} - H_k\Delta g^{(k)})}, \quad b^{(k)} = -\frac{H_k\Delta g^{(k)}}{\Delta g^{(k)\top}(\Delta x^{(k)} - H_k\Delta g^{(k)})}.$$

DFP Algorithm:
By Theorem 11.3, in the DFP algorithm applied to the quadratic with Hessian $Q = Q^\top$, we have

$$H_{k+1}\Delta g^{(k)} = \Delta x^{(k)}, \quad 0 \leq k.$$

Thus, conditions 1 and 2 are satisfied.
For condition 3:

$$U_k = H_{k+1} - H_k = \frac{\Delta x^{(k)}(\Delta x^{(k)})^\top}{\Delta x^{(k)\top}\Delta g^{(k)}} - \frac{H_k\Delta g^{(k)}[H_k\Delta g^{(k)}]^\top}{\Delta g^{(k)\top}H_k\Delta g^{(k)}}.$$

From this:

$$a^{(k)} = \frac{\Delta x^{(k)}}{\Delta x^{(k)\top}\Delta g^{(k)}}, \quad b^{(k)} = -\frac{H_k\Delta g^{(k)}}{\Delta g^{(k)\top}H_k\Delta g^{(k)}}.$$

BFGS Algorithm:
Recall that the BFGS algorithm is a special case of the DFP algorithm, it's a mirror or dual.

$$\begin{aligned} H_{k+1}^{\text{BFGS}} &= (B_{k+1})^{-1} \\ B_{k+1} &= \Delta g^{(i)} \Delta(x^{(i)})^{-1} \\ H_{k+1}^{\text{BFGS}} &= (B_{k+1})^{-1} = (\Delta g^{(i)}(\Delta x^{(i)})^{-1})^{-1} \\ &= \Delta x^{i}(\Delta g^{(i)})^{-1} \\ H_{k+1}^{\text{BFGS}} \Delta g^{(i)} &= \Delta x^{(i)} \end{aligned}$$

So, conditions 1 and 2 are satisfied.
The correction term update for the BFGS algorithm is given by

$$H_{k+1}^{\text{BFGS}} = H_k + \left(1 + \frac{\Delta g^{(k)\top} H_k \Delta g^{(k)}}{\Delta g^{(k)\top} \Delta x^{(k)}}\right) \frac{\Delta x^{(k)} (\Delta x^{(k)})^\top}{\Delta x^{(k)\top} \Delta g^{(k)}} - \frac{H_k \Delta g^{(k)} (\Delta x^{(k)})^\top + (H_k \Delta g^{(k)} (\Delta x^{(k)})^\top)^\top}{\Delta g^{(k)\top} \Delta x^{(k)}}.$$

From this, we can find that

$$a^{(k)} = \left(1 + \frac{\Delta g^{(k)\top} H_k \Delta g^{(k)}}{\Delta g^{(k)\top} \Delta x^{(k)}}\right) \frac{\Delta x^{(k)}}{\Delta x^{(k)\top} \Delta g^{(k)}} - \frac{H_k \Delta g^{(k)}}{\Delta g^{(k)\top} \Delta x^{(k)}}, \quad b^{(k)} = \frac{\Delta x^{(k)}}{\Delta g^{(k)\top} \Delta x^{(k)}}$$

5. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, consider an algorithm

$$x^{(k+1)} = x^{(k)} - \alpha_k H_k g^{(k)}$$

for finding the minimizer of f, where $g^{(k)} = \nabla f(x^{(k)})$ and $H_k \in \mathbb{R}^{n \times n}$ is symmetric. Suppose that

$$H_k = \phi H_k^{\text{DFP}} + (1 - \phi) H_k^{\text{BFGS}},$$

where $\phi \in \mathbb{R}$, and H_k^{DFP} and H_k^{BFGS} are matrices generated by the DFP and BFGS algorithms, respectively.

(a) Show that the algorithm above is a quasi-Newton algorithm. Is the above algorithm a conjugate direction algorithm?

(b) Suppose that $0 \leq \phi \leq 1$. Show that if $H_0^{\text{DFP}} > 0$ and $H_0^{\text{BFGS}} > 0$, then $H_k > 0$ for all k. What can you conclude from this about whether or not the algorithm has the descent property?

(a) To show the algorithm is a quasi-Newton algorithm, we will show

$$H_k \Delta g^{(i)} = \Delta x^{(i)}$$

$$H_k \Delta g^{(i)} = (\phi H_k^{\text{DFP}} + (1 - \phi) H_k^{\text{BFGS}}) \Delta g^{(i)}$$

Since DFP and BFGS are quasi-Newton algorithms, we know that the above relation holds, so distributing $\Delta g^{(i)}$

$$\begin{aligned} H_k \Delta g^{(i)} &= \phi \Delta x^{(i)} + (1 - \phi) \Delta x^{(i)} \\ &= \Delta x^{(i)} \end{aligned}$$

Since the above algorithm is a quasi-Newton algorithm, the algorithm is a conjugate direction algorithm. (Theorem 11.1)

(b) Observe that

$$(1 - \phi) \geq 0$$

Since $H_0^{\mathrm{DFP}} > 0$ and $H_0^{\mathrm{BFGS}} > 0$, it follows that $H_k > 0$ for all k. H_k is positive definite and symmetric. So, by Proposition 11.1, the algorithm has the descent property.