Math 164 Homework 10

Nick Monozon

Question 11.1

(a) For ease of notation, let
$$\phi(\alpha) = f(x^{(k)} + \alpha d^{(k)}).$$
By applying the Chain Rule, we get that
$$\phi'(\alpha) = d^{(k)\top} \nabla f(x^{(k)} + \alpha d^{(k)}).$$
So then
$$\phi'(0) = d^{(k)\top} g^{(k)}.$$
Since $\phi'(\alpha)$ is continuous, then if $d^{(k)\top} g^{(k)} < 0$, there exists some $\beta > 0$ such that for all $\alpha \in (0, \beta]$, $\phi(\alpha) < \phi(0)$. Recall how we defined $\phi(\alpha)$. This hence means that
$$f(x^{(k)} + \alpha d^{(k)}) < f(x^{(k)})$$
as desired.

(b) By part (a), since $\phi(\alpha) < \phi(0)$ for all $\alpha \in (0, \beta]$, we have that
$$\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} \phi(\alpha) \neq 0.$$
This necessarily implies that $a_k > 0$, so we have the intended result.

(c) We have that
$$d^{(k)\top} g^{(k+1)} = d^{(k)\top} \nabla f(x^{(k)} + \alpha_k d^{(k)}) = \phi_k'(\alpha_k).$$
From part (b), we showed that $a_k > 0$, so $\phi_k'(\alpha_k) > 0$. As such, we have that $g^{(k+1)\top} d^{(k)} = 0$.

(d)

1. For the steepest descent algorithm, we have that $d^{(k)} = -g^{(k)}$. As such, $d^{(k)\top} g^{(k)} = -\|g^{(k)}\|^2$. If $g^{(k)} \neq 0$, then $\|g^{(k)}\|^2 > 0$ and therefore $d^{(k)\top} g^{(k)} < 0$.
2. Next, for Newton's Method, assuming that the Hessian is positive definite, we have $d^{(k)} = -F(x^{(k)})^{-1} g^{(k)}$. Since $F(x^{(k)}) > 0$ by assumption, we also have that $F(x^{(k)})^{-1} > 0$. As such, given $g^{(k)} \neq 0$
$$d^{(k)\top} g^{(k)} = -g^{(k)\top} F(x^{(k)})^{-1} g^{(k)} < 0.$$
3. Now, for the conjugate gradient algorithm, we have that
$$d^{(k)} = -g^{(k)} + \beta_{k-1} d^{(k-1)}.$$
As such, it follows that
$$d^{(k)\top} g^{(k)} = -\|g^{(k)}\|^2 + \beta_{k-1} d^{(k-1)\top} g^{(k)}.$$
Using the result of part (c), we know that $g^{(k+1)\top} d^{(k)} = 0$. This means that
$$d^{(k)\top} g^{(k)} = -\|g^{(k)}\|^2 < 0.$$
4. Lastly, for the quasi-Newton algorithm, we have that $d^{(k)} = -H_k g^{(k)}$. Assuming that $H_k > 0$ and $g^{(k)} \neq 0$, we have that $d^{(k)\top} g^{(k)} = -g^{(k)\top} H_k g^{(k)} < 0$.

(e) Since $f(x)$ is quadratic, we have that $\nabla f(x) = Qx - b$. So we have that

$$\begin{aligned} d^{(k)\top} g^{(k+1)} &= d^{(k)\top}(Qx^{(k+1)} - b) \\ &= d^{(k)\top}(Q(x^{(k)} + \alpha_k d^{(k)}) - b) \\ &= a_k d^{(k)\top} Q d^{(k)} + d^{(k)\top}(Qx^{(k)} - b) \\ &= \alpha_k d^{(k)\top} Q d^{(k)} + d^{(k)\top} g^{(k)}. \end{aligned}$$

Using the result of part (c), we have that $g^{(k+1)\top} d^{(k)} = 0$. Hence, we can rearrange the expression and solve for α_k to obtain

$$\alpha_k = -\frac{d^{(k)\top} g^{(k)}}{d^{(k)\top} Q d^{(k)}}.$$

Question 11.3

(a) To begin, we have that

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

Applying the Chain Rule yields

$$\frac{d}{d\alpha} f(x^{(k)} + \alpha d^{(k)}) = (x^{(k)} + \alpha d^{(k)})^\top Q d^{(k)} - d^{(k)\top} b.$$

Letting $\frac{df}{d\alpha} = 0$, we write that

$$(x^{(k)\top} Q - b^\top) d^{(k)} = -\alpha d^{(k)\top} Q d^{(k)}.$$

Note that $g^{(k)} = x^{(k)\top} Q - b^\top$ and $d^{(k)\top} Q d^{(k)} > 0$ when $g^{(k)} \neq 0$. Hence, we rearrange the expression to obtain

$$\alpha_k = -\frac{g^{(k)\top} d^{(k)}}{d^{(k)\top} Q d^{(k)}} = \frac{g^{(k)\top} H_k g^{(k)}}{d^{(k)\top} Q d^{(k)}}.$$

(b) Since Q is a symmetric and positive definite matrix, we have that $\alpha_k > 0$ for $H_k = H_k^\top > 0$.

Question 11.4

(a) Given the algorithm

$$x^{(k+1)} = x^{(k)} - H g^{(k)},$$

we choose $H = F(x^{(k)})^{-1}$ to guarantee that given $x^{(0)}$ sufficiently close to x^*, $x^{(k)} \to x^*$ with order of convergence at least 2. This is precisely Newton's Method, which by Theorem 9.1 has order of convergence at least 2.

(b) Yes, this is a quasi-Newton method given the usual step size. If we apply the algorithm for the choice of H in part (a), then when applied to a quadratic function with Hessian matrix Q, we simply have that $H = Q^{-1}$, which satisfies the quasi-Newton condition.

Question 11.5

We want to minimize the quadratic function

$$f(x) = \frac{1}{2} x^\top \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x - x^\top \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 7$$

with the initial point $x^{(0)} = 0$ using the rank one correction algorithm. The gradient of f evaluated at $x^{(0)}$ is

$$\nabla f(x^{(0)}) = g^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

This vector is non-zero, so we proceed with the first iteration. Let $H_0 = I_2$. We have that

$$d^{(0)} = -H_0 g^{(0)} = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Using the result from 11.1(e), the step size can be calculated as

$$\alpha_0 = -\frac{d^{(0)\top} g^{(0)}}{d^{(0)\top} Q d^{(0)}} = -\frac{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}} = \frac{2}{3}.$$

And so then

$$x^{(1)} = x^{(0)} + \alpha_0 d^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix}.$$

The gradient of f evaluated at $x^{(1)}$ is

$$\nabla f(x^{(1)}) = g^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1/3 \\ -1/3 \end{bmatrix}.$$

This vector is non-zero, so we proceed with the second iteration. We compute H_1 using the formula

$$H_1 = H_0 + \frac{(\Delta x^{(0)} - H_0 \Delta g^{(0)})(\Delta x^{(0)} - H_0 \Delta g^{(0)})^\top}{\Delta g^{(0)\top} (\Delta x^{(0)} - H_0 \Delta g^{(0)})}.$$

We compute

$$\Delta x^{(0)} = x^{(1)} - x^{(0)} = \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix}$$

and

$$\Delta g^{(0)} = g^{(1)} - g^{(0)} = \begin{bmatrix} -1/3 \\ -1/3 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -4/3 \end{bmatrix}.$$

We also compute

$$\Delta x^{(0)} - H_0 \Delta g^{(0)} = \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2/3 \\ -4/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \end{bmatrix}$$

and

$$\Delta g^{(0)\top} (\Delta x^{(0)} - H_0 \Delta g^{(0)}) = \begin{bmatrix} 2/3 \\ -4/3 \end{bmatrix}^\top \begin{bmatrix} 0 \\ 2/3 \end{bmatrix} = -\frac{8}{9}.$$

Putting this all together,

$$H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} 0 & 0 \\ 0 & 4/9 \end{bmatrix}}{-8/9} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

So we have that

$$d^{(1)} = -H_1 g^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} -1/3 \\ -1/3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/6 \end{bmatrix}.$$

We compute that

$$\alpha_1 = -\frac{d^{(1)\top} g^{(1)}}{d^{(1)\top} Q d^{(1)}} = -\frac{\begin{bmatrix} 1/3 & 1/6 \end{bmatrix} \begin{bmatrix} -1/3 \\ -1/3 \end{bmatrix}}{\begin{bmatrix} 1/3 & 1/6 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/6 \end{bmatrix}} = 1.$$

So

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix} 2/3 \\ -2/3 \end{bmatrix} + \begin{bmatrix} 1/3 \\ 1/6 \end{bmatrix} = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}.$$

We expect that $x^* = x^{(2)}$. We see that

$$g^{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

So $x^* = x^{(2)} = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$ using the rank one correction algorithm.

Question 11.12

By Lemma 11.1, we have that

$$(A + uv^\top)^{-1} = A^{-1} - \frac{(A^{-1}u)(v^\top A^{-1})}{1 + v^\top A^{-1}u}$$

where $1 + v^\top A^{-1} u \neq 0$. Using the definitions from the hint, we can represent B_{k+1} as

$$\begin{aligned} B_{k+1} &= A_0 + u_0 v_0^\top + u_1 v_1^\top \\ &= A_1 + u_1 v_1^\top. \end{aligned}$$

Applying Lemma 11.1 yields

$$\begin{aligned} H_{k+1}^{BFGS} &= (B_{k+1})^{-1} \\ &= (A_1 + u_1 v_1^\top)^{-1} \\ &= A_1^{-1} - \frac{A_1^{-1} u_1 v_1^\top A_1^{-1}}{1 + v_1^\top A_1^{-1} u_1}. \end{aligned}$$

Substituting the above gives

$$B_{k+1}^{BFGS} = A_0^{-1} - \frac{A_0^{-1} u_0 v_0^\top A_0^{-1}}{1 + v_0^\top A_0^{-1} u_0} - \frac{\left(A_0^{-1} - \frac{A_0^{-1} u_0 v_0^\top A_0^{-1}}{1 + v_0^\top A_0^{-1} u_0}\right) u_1 v_1^\top \left(A_0^{-1} - \frac{A_0^{-1} u_0 v_0^\top A_0^{-1}}{1 + v_0^\top A_0^{-1} u_0}\right)}{1 + v_1^\top \left(A_0^{-1} - \frac{A_0^{-1} u_0 v_0^\top A_0^{-1}}{1 + v_0^\top A_0^{-1} u_0}\right) u_1}.$$

At this point, notice that $A_0 = B_k$. As such, $A_0^{-1} = B_k^{-1} = H_k$. Using this fact, the notation presented in the hint, the fact that because $H_k = B_k^{-1}$ we have

$$H_k B_k = B_k H_k = I_n,$$

and cancelling terms, we obtain

$$\begin{aligned} H_{k+1}^{BFGS} &= H_k - \frac{H_k \Delta g^{(k)} \Delta g^{(k)\top} H_k}{g^{(k)\top} \Delta x^{(k)} + \Delta g^{(k)\top} H_k \Delta g^{(k)}} \\ &- \frac{\left(\frac{H_k \Delta g^{(k)} \Delta g^{(k)\top} \Delta x^{(k)}}{g^{(k)\top} \Delta x^{(k)} + \Delta g^{(k)\top} H_k \Delta g^{(k)}} - \Delta x^{(k)}\right)\left(\Delta x^{(k)\top} - \frac{\Delta x^{(k)\top} \Delta g^{(k)} \Delta g^{(k)\top} H_k}{\Delta g^{(k)\top} \Delta x^{(k)} + \Delta g^{(k)\top} H_k \Delta g^{(k)}}\right)}{\Delta x^{(k)\top} B_k \Delta x^{(k)} - \Delta x^{(k)\top} B_k \Delta x^{(k)} + \frac{\Delta x^{(k)\top} \Delta g^{(k)} \Delta g^{(k)\top} \Delta x^{(k)}}{\Delta g^{(k)\top} \Delta x^{(k)} + \Delta g^{(k)\top} H_k \Delta g^{(k)}}}. \end{aligned}$$

Simplifying the expression and omitting some steps for brevity, we obtain

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

to be the BFGS update formula we desire.

Question 11.14

(a) Suppose that conditions 1-3 hold when applied to a quadratic. We want to show that for each $k = 0, \ldots, n-1$ and $i = 0, \ldots, k$ that $H_{k+1} \Delta g^{(i)} = \Delta x^{(i)}$. For $i = k$, it follows from conditions 1 and 2 that

$$\begin{aligned} H_{k+1} \Delta g^{(k)} &= H_k \Delta g^{(k)} + U_k \Delta g^{(k)} \\ &= H_k \Delta g^{(k)} + \Delta x^{(k)} - H_k \Delta g^{(k)} \\ &= \Delta x^{(k)} \end{aligned}$$

as desired. For the inductive step, we will use induction on k where $i = 0, \ldots, k-1$. Suppose that the proposition holds for $k-1$. We want to show that the result holds for k, so we fix i such that $0 \le i \le k-1$. We have that

$$\begin{aligned} H_{k+1} \Delta g^{(i)} &= H_k \Delta g^{(i)} + U_k \Delta g^{(i)} \\ &= \Delta x^{(i)} + U_k \Delta g^{(i)} \\ &= \Delta x^{(i)} + a^{(k)} \Delta x^{(k)\top} \Delta g^{(i)} + b^{(k)} \Delta g^{(k)\top} H_k \Delta g^{(i)}. \end{aligned}$$

We want to show that the last two terms are zero. For the middle term,

$$\begin{aligned}\Delta x^{(k)\top}\Delta g^{(i)} &= \Delta x^{(k)\top} Q\Delta x^{(i)} \\ &= \alpha_k\alpha_i d^{(k)\top} Q d^{(i)} \\ &= 0\end{aligned}$$

by the inductive hypothesis, which implies Q-conjugacy. For the last term, we have that

$$\begin{aligned}\Delta g^{(k)\top} H_k \Delta g^{(i)} &= \Delta g^{(k)\top}\Delta x^{(i)} \\ &= \Delta x^{(k)\top} Q\Delta x^{(i)} \\ &= \alpha_k\alpha_i d^{(k)\top} Q d^{(i)} \\ &= 0\end{aligned}$$

by the inductive hypothesis, which implies Q-conjugacy. The algorithm is quasi-Newton, so this completes the proof.

(b) The rank-one, DFP, and BFGS algorithms all satisfy the three conditions in part (a) when applied to a quadratic. For the rank-one algorithm, we have that

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

For the DFP algorithm, we have that

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

And for the BFGS algorithm we have that

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

Question 11.15

(a) Suppose that we apply the algorithm to a quadratic function. Then, by the quasi-Newton property of the DFP algorithm, we have that $H_{k+1}^{DFP} = \Delta g^{(i)}$ for each $i = 0, 1, \dots, k$. This property holds similarly for the BFGS algorithm. This means that for a given H_k, for each $i = 0, 1, \dots, k$,

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

As such, we have that the algorithm is quasi-Newton and therefore also a conjugate direction algorithm.

(b) By Theorem 11.4 and the properties of the BFGS algorithm, we have that $H_k^{BFGS} > 0$ and $H_k^{DFP} > 0$. Therefore, for $x \neq 0$,

$$x^\top H_k x = \phi x^\top H_k^{DFP} x + (1-\phi) x^\top H_k^{BFGS} x > 0.$$

where $\phi, 1-\phi \geq 0$ as $0 \leq \phi \leq 1$. As such, $H_k > 0$, so by Proposition 11.1 the algorithm has the descent property if α_k is computed using line search.

Question 11.16

Suppose that for a quadratic form with Hessian Q, $H_{k+1}\Delta g^{(i)} = \rho_i \Delta x^{(i)}$ for $0 \leq i \leq k, k \leq n-1$. We want to prove that if $\alpha_i \neq 0$ for each $i = 0, 1, \dots, k$, then $d^{(0)}, \dots, d^{(k+1)}$ are Q-conjugate. We will proceed by induction. We consider the base case where $k = 0$, so we want to show that $d^{(0)}$ and $d^{(1)}$ are Q-conjugate. Since $\alpha_0 \neq 0$, we write that $d^{(0)} = \Delta x^{(0)}/\alpha_0$. It follows that

$$\begin{aligned} d^{(1)\top} Q d^{(0)} &= -g^{(1)\top} H_1 Q d^{(0)} \\ &= -g^{(1)\top} H_1 \frac{Q\Delta x^{(0)}}{\alpha_0} \\ &= -g^{(1)\top} \frac{H_1 \Delta g^{(0)}}{\alpha_0} \\ &= -g^{(1)\top} \frac{\rho_0 \Delta x^{(0)}}{\alpha_0} \\ &= -\rho_0 g^{(1)\top} d^{(0)}. \end{aligned}$$

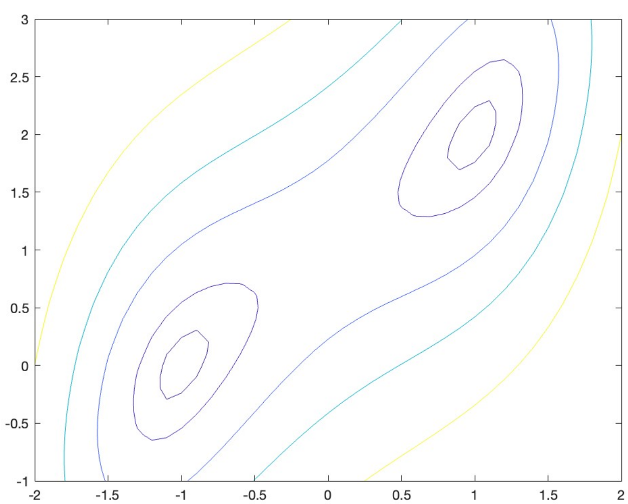
However, since $\alpha_0 > 0$ is the minimizer of $\phi(\alpha) = f(x^{(0)} + \alpha d^{(0)})$, we have that $g^{(1)\top} d^{(0)} = 0$. Therefore, $d^{(1)\top} Q d^{(0)} = 0$, so the base case holds. We will now proceed to the inductive step. Assume that the proposition holds for all $k-1$ where $k < n-1$. We want to show that it also holds for k, which means that $d^{(0)}, \ldots, d^{(k+1)}$ are Q-conjugate. As such, for each i such that $0 \leq i \leq k$ where $\alpha_i \neq 0$, we have that

$$\begin{aligned} d^{(k+1)\top} Q d^{(i)} &= -g^{(k+1)\top} H_{k+1} Q d^{(i)} \\ &= -g^{(k+1)\top} H_{k+1} \frac{Q\Delta x^{(i)}}{\alpha_i} \\ &= -g^{(k+1)\top} \frac{H_{k+1} \Delta g^{(i)}}{\alpha_i} \\ &= -g^{(k+1)\top} \frac{\rho_i \Delta x^{(i)}}{\alpha_i} \\ &= -\rho_i g^{(k+1)\top} d^{(i)}. \end{aligned}$$

Since we assumed that $d^{(0)}, \ldots, d^{(k)}$ are Q-conjugate, we have by Lemma 10.2 that $g^{(k+1)\top} d^{(i)} = 0$. As such, $d^{(k+1)\top} Q d^{(i)} = 0$, so $d^{(0)}, \ldots, d^{(k+1)}$ are Q-conjugate, and we're done.

Question 11.18

(a) The levels set of f at the specified levels (-0.72, -0.6, -0.2, 0.5, 2) is below.



(b) Using a script I found online for the DFP algorithm, the initial point $[0,0]^\top$ converges to $[-1,0]^\top$ while the algorithm converges to $[1,2]^\top$ for the initial point $[1.5,1]^\top$. These points satisfy the SOSC, so they are both local minimizers. It seems that in this case, the DFP algorithm converges to the local minimizer nearest to the initial point.