1 Question 1

Assume that $f : D \to R$ is twice continuously differentiable for all $x \in D$, where the domain D of f is an open, convex subset of $\mathbb{R}^n$. Show that, its Hessian matrix, $\nabla^2 f(x)$, is symmetric positive-semi-definite for all $x \in D$ if and only if f is a convex function on D.
Moreover, show that, if its Hessian matrix, $\nabla^2 f(x)$, is symmetric positive-definite for all $x \in D$, then f is a strictly convex function on D.
Show that the converse of this last statement is not true. That is, there is a strictly convex function on an open, convex domain D such that its Hessian matrix, $\nabla^2 f(x)$, is not symmetric positive-definite for all $x \in D$.

$''\Rightarrow''$ Assume the Hessian matrix, $\nabla^2 f(x)$, is symmetric positive-semi-definite for all $x \in D$
Let $x, y \in D$ and let $\alpha \in [0,1]$
Let $z, w \in D$
let $p = w - z$. Since $f : D \to R$ is twice continuously differentiable, we can apply Taylor Theorem from the text book
$f(z+p) = f(z) + \nabla f(z)^T p + \frac{1}{2} p^T \nabla^2 f(z+tp)p$ for some $t \in (0,1)$
but since $\nabla^2 f(x)$ is symmetric positive-semi-definite everywhere, we have that $p^T \nabla^2 f(z+tp)p \geq 0$, so
$f(z+p) \geq f(z) + \nabla f(z)^T p$

$$f(w) \geq f(z) + \nabla f(z)^T (w - z)$$

Since z and w where arbitrary, it is true for $w = x$, $z = \alpha x + (1-\alpha)y$ and $w = y$, $z = \alpha x + (1-\alpha)y$. We have

$$f(x) \geq f(\alpha x + (1-\alpha)y) + \nabla f(\alpha x + (1-\alpha)y)^T (x - \alpha x - (1-\alpha)y) \qquad \text{(i)}$$
$$f(y) \geq f(\alpha x + (1-\alpha)y) + \nabla f(\alpha x + (1-\alpha)y)^T (y - \alpha x - (1-\alpha)y) \qquad \text{(ii)}$$

Manipulate (i)

$$\begin{aligned}
f(x) &\geq f(\alpha x + (1-\alpha)y) + \nabla f(\alpha x + (1-\alpha)y)^T ((1-\alpha)x - (1-\alpha)y) && \text{simplified} \\
f(x) &\geq f(\alpha x + (1-\alpha)y) + (1-\alpha)\nabla f(\alpha x + (1-\alpha)y)^T (x - y) && \text{factor out } (1-\alpha) \\
\alpha f(x) &\geq \alpha f(\alpha x + (1-\alpha)y) + \alpha(1-\alpha)\nabla f(\alpha x + (1-\alpha)y)^T (x - y) && \text{multiply both sides by } \alpha
\end{aligned}$$

Manipulate (ii)

$$\begin{aligned}
f(y) &\geq f(\alpha x + (1-\alpha)y) + \nabla f(\alpha x + (1-\alpha)y)^T (\alpha y - \alpha x) && \text{simplified} \\
f(y) &\geq f(\alpha x + (1-\alpha)y) + \alpha\nabla f(\alpha x + (1-\alpha)y)^T (y - x) && \text{factor out } \alpha \\
(1-\alpha)f(y) &\geq (1-\alpha)f(\alpha x + (1-\alpha)y) + \alpha(1-\alpha)\nabla f(\alpha x + (1-\alpha)y)^T (y - x) && \text{multiply both sides by } (1-\alpha) \\
(1-\alpha)f(y) &\geq (1-\alpha)f(\alpha x + (1-\alpha)y) - \alpha(1-\alpha)\nabla f(\alpha x + (1-\alpha)y)^T (x - y) && \text{factor out } -1
\end{aligned}$$

Add the two manipulated inequalities together.

$$\begin{aligned}
\alpha f(x) + (1-\alpha)f(y) &\geq \alpha f(\alpha x + (1-\alpha)y) + (1-\alpha)f(\alpha x + (1-\alpha)y) \\
\alpha f(x) + (1-\alpha)f(y) &\geq f(\alpha x + (1-\alpha)y)
\end{aligned}$$

Since x, y and $\alpha \in [0,1]$ were arbitrary, we have that $\forall x, y \in D$ and any $\alpha \in [0,1]$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$

Finally because the domain D of f is convex, f is a convex function on D.

$''\Leftarrow''$ Assume f is a convex function on D. By contradiction
Assume there exists a $z \in D$ s.t. $\nabla^2 f(z)$ is not positive-semi-definite, so $\exists p \neq 0$ s.t. $p^T \nabla^2 f(z) p < 0$
$\nabla^2 f(x)$ is continious everywhere by assumption, specifically around z, so $\exists \alpha > 0$ s.t. $p^T \nabla^2 f(z + t\alpha p)p < 0$ for all $t \in (0,1)$
let $y = \alpha p + z$. By Taylor's Theorem
$f(z + \alpha p) = f(z) + \nabla f(z)^T (\alpha p) + \frac{\alpha^2}{2} p^T \nabla^2 f(z + t\alpha p)p$ for some $t \in (0,1)$
Because $\frac{\alpha^2}{2} p^T \nabla^2 f(z + t\alpha p)p < 0$
$f(z + \alpha p) < f(z) + \nabla f(z)^T (\alpha p)$

$$f(y) < f(z) + \nabla f(z)^T (y - z)$$

let $g = y - z$
$\nabla f(x)$ is continuous everywhere, specifically near z. So $\exists \beta > 0$, s.t $f(y) < f(z) + \nabla f(z + t\beta g)^T(y - z)$ for all $t \in (0, 1)$
For the above choose a small β s.t $\beta < 1$. Multiplying both sides by β and rearranging we get

$$\beta f(y) < \beta f(z) + \beta \nabla f(z + t\beta g)^T(y - z) \implies \beta \nabla f(z + t\beta g)^T(y - z) > \beta f(y) - \beta f(z) \quad \forall t \in (0, 1)$$

By Taylor's Theorem

$$\begin{aligned}
f(z + \beta g) &= f(z) + \nabla f(z + t\beta g)^T(\beta g) && \text{for some } t \in (0, 1) \\
f(z + \beta(y - z)) &= f(z) + \beta \nabla f(z + t\beta g)^T(y - z) && \text{since } g = y - z \\
f(z + \beta(y - z)) &> f(z) + \beta f(y) - \beta f(z) && \text{by above, holds for all t, specifically for this t} \\
f(z - \beta z + \beta y) &> f(z) - \beta f(z) + \beta f(y) \\
f((1 - \beta)z + \beta y) &> (1 - \beta)f(z) + \beta f(y)
\end{aligned}$$

Which is a contradiction,
By our assumption of f being convex and choice of β, we have that

$$f((1 - \beta)z + \beta y) \leq (1 - \beta)f(z) + \beta f(y)$$

Hence $\nabla^2 f(x)$ is symmetric positive-semi-definite for all $x \in D$

Moreover, show that, if its Hessian matrix, $\nabla^2 f(x)$, is symmetric positive-definite for all $x \in D$, then f is a strictly convex function on D.

"$\Rightarrow$" Assume the Hessian matrix, $\nabla^2 f(x)$, is symmetric positive-definite for all $x \in D$
Let $x, y \in D$, $x \neq y$ and let $\alpha \in (0, 1)$
Let $z, w \in D$ s.t. $z \neq w$
let $p = w - z$, note ($p \neq \mathbf{0}$). Since $f : D \to R$ is twice continuously differentiable, again can apply Taylor Theorem
$f(z + p) = f(z) + \nabla f(z)^T p + \frac{1}{2} p^T \nabla^2 f(z + tp)p$ for some $t \in (0, 1)$
but since $\nabla^2 f(x)$ is symmetric positive-definite everywhere and $p \neq \mathbf{0}$, we have that $p^T \nabla^2 f(z + tp)p > 0$, so
$f(z + p) > f(z) + \nabla f(z)^T p$

$$f(w) > f(z) + \nabla f(z)^T(w - z)$$

Since z and w where arbitrary, it is true for $w = x$, $z = \alpha x + (1 - \alpha)y$ and $w = y$, $z = \alpha x + (1 - \alpha)y$. Since $x \neq y$ and $\alpha \in (0, 1)$, $w \neq z$ in both cases so we have

$$\begin{aligned}
f(x) &> f(\alpha x + (1 - \alpha)y) + \nabla f(\alpha x + (1 - \alpha)y)^T(x - \alpha x - (1 - \alpha)y) && \text{(i)} \\
f(y) &> f(\alpha x + (1 - \alpha)y) + \nabla f(\alpha x + (1 - \alpha)y)^T(y - \alpha x - (1 - \alpha)y) && \text{(ii)}
\end{aligned}$$

Manipulate (i)

$$\begin{aligned}
f(x) &> f(\alpha x + (1 - \alpha)y) + \nabla f(\alpha x + (1 - \alpha)y)^T((1 - \alpha)x - (1 - \alpha)y) && \text{simplified} \\
f(x) &> f(\alpha x + (1 - \alpha)y) + (1 - \alpha)\nabla f(\alpha x + (1 - \alpha)y)^T(x - y) && \text{factor out } (1 - \alpha) \\
\alpha f(x) &> \alpha f(\alpha x + (1 - \alpha)y) + \alpha(1 - \alpha)\nabla f(\alpha x + (1 - \alpha)y)^T(x - y) && \text{multiply both sides by } \alpha
\end{aligned}$$

Manipulate (ii)

$$\begin{aligned}
f(y) &> f(\alpha x + (1 - \alpha)y) + \nabla f(\alpha x + (1 - \alpha)y)^T(\alpha y - \alpha x) && \text{simplified} \\
f(y) &> f(\alpha x + (1 - \alpha)y) + \alpha \nabla f(\alpha x + (1 - \alpha)y)^T(y - x) && \text{factor out } \alpha \\
(1 - \alpha)f(y) &> (1 - \alpha)f(\alpha x + (1 - \alpha)y) + \alpha(1 - \alpha)\nabla f(\alpha x + (1 - \alpha)y)^T(y - x) && \text{multiply both sides by } (1 - \alpha) \\
(1 - \alpha)f(y) &> (1 - \alpha)f(\alpha x + (1 - \alpha)y) - \alpha(1 - \alpha)\nabla f(\alpha x + (1 - \alpha)y)^T(x - y) && \text{factor out } -1
\end{aligned}$$

Add the two manipulated inequalities together.

$$\begin{aligned}
\alpha f(x) + (1 - \alpha)f(y) &> \alpha f(\alpha x + (1 - \alpha)y) + (1 - \alpha)f(\alpha x + (1 - \alpha)y) \\
\alpha f(x) + (1 - \alpha)f(y) &> f(\alpha x + (1 - \alpha)y)
\end{aligned}$$

Since $x, y, x \neq y$ and $\alpha \in (0, 1)$ were arbitrary, we have that $\forall x, y \in D$, $x \neq y$ and any $\alpha \in (0, 1)$

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

Finally because the domain D of f is convex, f is a strictly convex function on D.

Show that the converse of this last statement is not true. That is, there is a strictly convex function on an open, convex domain D such that its Hessian matrix, $\nabla^2 f(x)$, is not symmetric positive-definite for all $x \in D$.

Take the domain D to be $\mathbf{R}$, which is a vector space and so is a convex set.
Take the function to be $f(x) = x^4$
First can show $g(x) = x^2$ is a strictly convex function.
Let $x, y \in \mathbf{R}$, $x \neq y$ and $\alpha \in (0,1)$

$$\begin{aligned}
g(\alpha x + (1-\alpha)y) &= (\alpha x + (1-\alpha)y)^2 \\
&= \alpha^2 x^2 + 2\alpha(1-\alpha)xy + (1-\alpha)^2 y^2 \\
&= \alpha^2 x^2 - \alpha x^2 + 2\alpha(1-\alpha)xy + (1-\alpha)^2 y^2 - (1-\alpha)y^2 + \alpha x^2 + (1-\alpha)y^2 \\
&= \alpha(\alpha-1)x^2 + 2\alpha(1-\alpha)xy + (1-\alpha)((1-\alpha)-1)y^2 + \alpha x^2 + (1-\alpha)y^2 \\
&= -\alpha(1-\alpha)x^2 + 2\alpha(1-\alpha)xy - \alpha(1-\alpha)y^2 + \alpha x^2 + (1-\alpha)y^2 \\
&= -\alpha(1-\alpha)(x^2 - 2xy + y^2) + \alpha x^2 + (1-\alpha)y^2 \\
&= -\alpha(1-\alpha)(x-y)^2 + \alpha x^2 + (1-\alpha)y^2 \\
&< \alpha x^2 + (1-\alpha)y^2 = \alpha g(x) + (1-\alpha)g(y) && \text{since } \alpha(1-\alpha)(x-y)^2 > 0
\end{aligned}$$

$$\begin{aligned}
f(\alpha x + (1-\alpha)y) &= (\alpha x + (1-\alpha)y)^4 \\
&= ((\alpha x + (1-\alpha)y)^2)^2 \\
&< ((\alpha x^2 + (1-\alpha)y^2)^2 && \text{by above argument} \\
&= ((\alpha s + (1-\alpha)t)^2 && \text{let } s = x^2 \text{ and } t = y^2 \\
&< \alpha s^2 + (1-\alpha)t^2 && \text{by above argument} \\
&= \alpha(x^2)^2 + (1-\alpha)(y^2)^2 \\
&= \alpha x^4 + (1-\alpha)y^4 \\
&= \alpha f(x) + (1-\alpha)f(y)
\end{aligned}$$

So since for arbitrary x, y, $x \neq y$ and arbitrary $\alpha \in (0,1)$

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$$

Therefore the function $f(x) = x^4$ is a strictly convex function.
$\nabla f(x) = f'(x) = 4x^3$, $\nabla^2 f(x) = f''(x) = 12x^2$, $\nabla^2 f(x)$ being positive definite in 1D is $f''(x) > 0$
But look at $x^* = 0$, all minimum of the function. $\nabla^2 f(x^*) f''(x^*) = 0$, not positive definite
So there is a strictly convex function such that it's Hessian is not symmetric positive definite everywhere.

2 Question 2

Do question 2.1 on page 27 of your textbook
Compute the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ of the Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Show that $x^* = [1,1]^T$ is the only local minimizer of this function, and that the Hessian matrix at that point is positive definite.

$$\nabla f(x) = \begin{bmatrix} 200(x_2 - x_1^2)(-2x_1) + 2(1-x_1)(-1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} -400x_1x_2 + 400x_1^3 - 2 + 2x_1 \\ 200x_2 - 200x_1^2 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

$$\nabla f(x^*) = \begin{bmatrix} -400(1)(1) + 400(1)^3 - 2 + 2(1) \\ 200(1) - 200(1)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \nabla^2 f(x^*) = \begin{bmatrix} -400(1) + 1200(1)^2 + 2 & -400(1) \\ -400(1) & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

We know that a symmetric matrix is positive definite if all of its eigenvalues are positive (proved in class). What are the eigenvalues of $\nabla^2 f(x^*)$

$$\begin{aligned}
det(\nabla^2 f(x^*) - \lambda I) &= 0 \\
(802-\lambda)(200-\lambda) - 400(400) &= 0 \\
160400 - 802\lambda - 200\lambda + \lambda^2 - 160000 &= 0 \\
\lambda^2 - 1002\lambda + 400 &= 0 \\
\lambda &= \frac{1002 \pm \sqrt{1002^2 - 4(400)}}{2} \\
\lambda_1 = 1001.6\ ,\lambda_2 = 0.399
\end{aligned}$$

Both eigenvalues are positive so $\nabla^2 f(x^*)$ is positive definite.
Now since $\nabla^2 f(x^*)$ is positive definite and $\nabla f(x^*) = 0$, by Theorem 2.4 (Nocedal textbook), x^* is a strict local minimizer of $f(x)$
Now to show that x^* is the only minimizer.
We know by Theorem 2.2 (Nocedal textbook) that if y is local minimizer, then $\nabla f(y) = 0$, so is $\nabla f(x) = 0$ at any other x?

$$\nabla f(x^*) = 0 \implies \begin{bmatrix} 200(x_2 - x_1^2)(-2x_1) + 2(1-x_1)(-1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} -400x_1x_2 + 400x_1^3 - 2 + 2x_1 = 0 \\ 200x_2 - 200x_1^2 = 0 \end{bmatrix}$$

$$200x_2 - 200x_1^2 = 0 \implies x_2 - x_1^2 = 0 \implies x_2 = x_1^2$$

Plugging this into the other equation.

$$-400x_1x_1^2 + 400x_1^3 - 2 + 2x_1 = 0 \implies -2 + 2x_1 = 0 \implies 2x_1 = 2 \implies x_1 = 1$$

Then $x_2 = x_1^2 = 1^2 = 1$, so the only point where the gradient is 0 is the point x^*, so x^* is the only local minimizer of $f(x)$

3 Question 3

Do question 2.8 on page 28 of your textbook Suppose that f is a convex function. Show that the set of global minimizers of f is a convex set.

Assume f is a convex function.
Let $F = \{x \mid f(x) \leq f(y) \ \ \forall y\}$
If f has no global minimizers, the set F is vacuously convex. We are done.
Else suppose F is nonempty, and so has at least one global minimizer.
Let $x, y \in F$ and let $\alpha \in [0,1]$
So $x \leq f(z)\ \forall z$ and $y \leq f(z)\ \forall z$ by definition of F. Therefore we have $f(x) \leq f(y)$ and $f(y) \leq f(x)$, implying that $f(x) = f(y)$

$$\begin{aligned}
f(\alpha x + (1-\alpha)y) &\leq \alpha f(x) + (1-\alpha) f(y) && \text{since } f \text{ is convex} \\
&= \alpha f(x) + (1-\alpha) f(x) && \text{since } f(x) = f(y) \\
&= f(x)
\end{aligned}$$

But $f(x) \leq f(z)\ \forall z$, so $f(x) \leq f(\alpha x + (1-\alpha)y)$
This and the inequality before it imply that $f(\alpha x + (1-\alpha)y) = f(x)$ and consequently $f(\alpha x + (1-\alpha)y) \leq f(z)\ \forall z$
Hence $f(\alpha x + (1-\alpha)y) \in F$
Since $x, y \in F$ and $\alpha \in [0,1]$ were arbitrary, we have

$$f(\alpha x + (1-\alpha)y) \in F \quad \forall x, y \in F, \forall \alpha \in [0,1]$$

Hence F, the set of global minimizers of f, is a convex set.

4 Question 4

Do question 2.11 on page 29 of your textbook
Show that the symmetric rank-one update (2.18) and the BFGS update (2.19) are scale-invariant if the initial Hessian approximations B_0 are chosen appropriately. That is, using the notation of the previous exercise, show that if these methods are applied to $f(x)$ starting from $x_0 = Sz_0 + s$ with initial Hessian B_0, and to $\hat{f}(z)$ starting from z_0 with initial Hessian $S^T B_0 S$, then all iterates are related by $x_k = Sz_k + s$. (Assume for simplicity that the methods take unit step lengths.)

Will use the facts from question 2.10 $\quad \nabla f(z) = S^T \nabla f(x) \quad \nabla^2 f(z) = S^T \nabla^2 f(x) S$

Let $P(k)$: At iteration k, $x_k = Sz_k + s$, and the Hessian for the two are related by B_k for $f(x)$ and $S^T B_k S$ for $f(z)$
Base Case: $P(0)$ is true by starting assumption of the question.
Induction Step: Assume that $P(k)$ holds, WTP that $P(k+1)$ holds.

$$\begin{aligned}
p_k^x &= -B_k^{-1} \nabla f(x_k) \\
p_k^z &= -(S^T B_k S)^{-1} \nabla f(z_k) \\
&= -S^{-1} B_k^{-1} (S^T)^{-1} \nabla f(z_k) && \text{by prop of inverse, } (AB)^{-1} = B^{-1}A^{-1} \\
&= -S^{-1} B_k^{-1} (S^T)^{-1} S^T \nabla f(x_k) && \text{by fact from 2.10} \\
&= -S^{-1} B_k^{-1} \nabla f(x_k) && \text{by def of inverse} \\
&= S^{-1} p_k^x
\end{aligned}$$

update rules $\quad x_{k+1} = x_k + p_k^x \quad z_{k+1} = z_k + p_k^z = z_k + S^{-1} p_k^x$

$$\begin{aligned}
Sz_{k+1} + s &= S(z_k + S^{-1} p_k^x) + s \\
&= Sz_k + p_k^x + s && \text{by distir. inverse} \\
&= x_k + p_k^x \\
&= x_{k+1}
\end{aligned}$$

So the first part of P(k+1) holds

$$\begin{aligned}
y_k^x &= \nabla f(x_{k+1}) - \nabla f(x_k) \\
y_k^z &= \nabla f(z_{k+1}) - \nabla f(z_k) \\
&= S^T \nabla f(x_{k+1}) - S^T \nabla f(x_k) && \text{by fact from 2.10} \\
&= S^T (\nabla f(x_{k+1}) - \nabla f(x_k)) \\
&= S^T y_k^x
\end{aligned}$$

SR1 Update for the Hessians

$$\begin{aligned}
B_{k+1}^x &= B_k + \frac{(y_k^x - B_k p_k^x)(y_k^x - B_k p_k^x)^T}{(y_k^x - B_k p_k^x)^T p_k^x} && \text{by update rule} \\
B_{k+1}^z &= S^T B_k S + \frac{(y_k^z - (S^T B_k S) p_k^z)(y_k^z - (S^T B_k S) p_k^z)^T}{(y_k^z - (S^T B_k S) p_k^z)^T p_k^z} && \text{by P(k) assumption relating Hessians} \\
&= S^T B_k S + \frac{((S^T y_k^x) - (S^T B_k S)(S^{-1} p_k^x))((S^T y_k^x) - (S^T B_k S)(S^{-1} p_k^x))^T}{((S^T y_k^x) - (S^T B_k S)(S^{-1} p_k^x))^T (S^{-1} p_k^x)} && \text{by } p_k^z = S^{-1} p_k^x \text{ and } y_k^z = S^T y_k^x \\
&= S^T B_k S + \frac{(S^T y_k^x - S^T B_k p_k^x)(S^T y_k^x - S^T B_k p_k^x)^T}{(S^T y_k^x - S^T B_k p_k^x)^T (S^{-1} p_k^x)} && \text{by inverses canceling} \\
&= S^T B_k S + \frac{S^T (y_k^x - B_k p_k^x)(S^T (y_k^x - B_k p_k^x))^T}{(S^T (y_k^x - B_k p_k^x))^T (S^{-1} p_k^x)} \\
&= S^T B_k S + \frac{S^T (y_k^x - B_k p_k^x)(y_k^x - B_k p_k^x)^T S}{(y_k^x - B_k p_k^x)^T S (S^{-1} p_k^x)} && (AB)^T = B^T A^T \text{ and } (A^T)^T = A \\
&= S^T B_k S + \frac{S^T (y_k^x - B_k p_k^x)(y_k^x - B_k p_k^x)^T S}{(y_k^x - B_k p_k^x)^T p_k^x} \\
&= S^T \left(B_k + \frac{(y_k^x - B_k p_k^x)(y_k^x - B_k p_k^x)^T}{(y_k^x - B_k p_k^x)^T p_k^x} \right) S = S^T B_{k+1}^x S
\end{aligned}$$

BFGS Update for the Hessians

$$
\begin{aligned}
B_{k+1}^x &= B_k - \frac{B_k p_k^x p_k^{xT} B_k}{p_k^{xT} B_k p_k^x} + \frac{y_k^x y_k^{xT}}{y_k^{xT} p_k^x} && \text{by update rule} \\
B_{k+1}^z &= S^T B_k S - \frac{(S^T B_k S) p_k^z p_k^{zT} (S^T B_k S)}{p_k^{zT} (S^T B_k S) p_k^z} + \frac{y_k^z y_k^{zT}}{y_k^{zT} p_k^z} && \text{Hessian is } S^T B_k S \\
&= S^T B_k S - \frac{(S^T B_k S)(S^{-1} p_k^x)(S^{-1} p_k^x)^T (S^T B_k S)}{(S^{-1} p_k^x)^T (S^T B_k S)(S^{-1} p_k^x)} + \frac{(S^T y_k^x)(S^T y_k^x)^T}{(S^T y_k^x)^T (S^{-1} p_k^x)} && p_k^z = S^{-1} p_k^x \text{ and } y_k^z = S^T y_k^x \\
&= S^T B_k S - \frac{S^T B_k S S^{-1} p_k^x p_k^{xT} (S^T)^{-1} S^T B_k S}{p_k^{xT} (S^T)^{-1} S^T B_k S S^{-1} p_k^x} + \frac{S^T y_k^x y_k^{xT} S}{y_k^{xT} S S^{-1} p_k^x} && (AB)^T = B^T A^T \text{ and } (A^{-1})^T = (A^T)^{-1} \\
&= S^T B_k S - \frac{S^T B_k p_k^x p_k^{xT} B_k S}{p_k^{xT} B_k p_k^x} + \frac{S^T y_k^x y_k^{xT} S}{y_k^{xT} p_k^x} && CA^{-1}A = CI = C \\
&= S^T \left(B_k - \frac{S^T B_k p_k^x p_k^{xT} B_k}{p_k^{xT} B_k p_k^x} + \frac{y_k^x y_k^{xT}}{y_k^{xT} p_k^x} \right) S \\
&= S^T B_{k+1}^x S
\end{aligned}
$$

So in either update, we have that the Hessians are still related by

$$B_{k+1}^z = S^T B_{k+1}^x S$$

and

$$x_{k+1} = S z_{k+1} + s$$

Therefore P(k+1) holds and so by induction $\forall k, P(k)$ holds.
All iterates are related by $x_k = S z_k + s$

5 Question 5

Do question 2.12 on page 29 of your textbook.
Suppose that a function f of two variables is poorly scaled at the solution x^* (local minimizer). Write two Taylor expansions of f around x^* — one along each coordinate direction—and use them to show that the Hessian $\nabla^2 f(x^*)$ is ill-conditioned.
If A is symmetric positive-semi-definite, then

$$cond_2(A) = \|A\|_2 \left\|A^{-1}\right\|_2 = \frac{\lambda_{max}}{\lambda_{min}}$$

Let $u = \alpha e_1$ and $v = \alpha e_2$ where $e_1 = [1,0]^T$ and $e_2 = [0,1]^T$ for some small α such that they satisfy

$$\left|f(x^* + u) - f(x^*)\right| >> \left|f(x^* + v) - f(x^*)\right|$$

The Taylor expansions along each direction

$$
\begin{aligned}
f(x^* + u) &= f(x^*) + \nabla f(x^*)^T u + \frac{1}{2} u^T \nabla^2 f(x^* + t_1 u) u && \text{for some } t_1 \in (0,1) \\
f(x^* + v) &= f(x^*) + \nabla f(x^*)^T v + \frac{1}{2} v^T \nabla^2 f(x^* + t_2 v) v && \text{for some } t_2 \in (0,1)
\end{aligned}
$$

$$
\begin{aligned}
\left|f(x^* + u) - f(x^*)\right| &>> \left|f(x^* + v) - f(x^*)\right| \\
\left|\nabla f(x^*)^T u + \frac{1}{2} u^T \nabla^2 f(x^* + t_1 u) u\right| &>> \left|\nabla f(x^*)^T v + \frac{1}{2} v^T \nabla^2 f(x^* + t_2 v) v\right| \\
\left|\frac{1}{2} u^T \nabla^2 f(x^* + t_1 u) u\right| &>> \left|\frac{1}{2} v^T \nabla^2 f(x^* + t_2 v) v\right| && \text{since } x^* \text{ is a minimizer} \\
\left|\frac{\alpha^2}{2} e_1^T \nabla^2 f(x^* + t_1 \alpha e_1) e_1\right| &>> \left|\frac{\alpha^2}{2} e_2^T \nabla^2 f(x^* + t_2 \alpha e_2) e_2\right| && \frac{\alpha^2}{2} \text{ is positive, can factor out} \\
\left|\frac{d^2 f}{dx_1^2}|_{x^* + t_1 \alpha e_1}\right| &>> \left|\frac{d^2 f}{dx_2^2}|_{x^* + t_2 \alpha e_2}\right| && e_1^T A e_1 \text{ extracts top left entry}
\end{aligned}
$$

Since α was small, the above is a good approximation of the entries of $\nabla^2 f(x^*)$

$$\nabla^2 f(x^*) = \begin{bmatrix} \frac{d^2 f}{dx_1^2}|_{x^*} & s \\ s & \frac{d^2 f}{dx_2^2}|_{x^*} \end{bmatrix} \qquad \text{where s is some real number (symmetric matrix)}$$

We know two eigenvalues exist and that they are both ≥ 0. Because matrix is symmetric positive definite.
Let $\lambda_{max} = max(\lambda_1, \lambda_2)$, $\lambda_{min} = min(\lambda_1, \lambda_2)$ and let $a = \frac{d^2 f}{dx_1^2}|_{x^*}$ and $b = \frac{d^2 f}{dx_2^2}|_{x^*}$
From the appendix of the Nocedal textbook we know.

$$Trace(\nabla^2 f(x^*)) = \lambda_{max} + \lambda_{min} \qquad det(\nabla^2 f(x^*)) = \lambda_{max}\lambda_{min}$$
$$det(\nabla^2 f(x^*)) = ab - s^2 = \lambda_{max}\lambda_{min} \geq 0$$

This means that a and b must both be positive (or both be negative).

$$Trace(\nabla^2 f(x^*)) = a + b = \lambda_{max} + \lambda_{min} \geq 0$$

So hence a and b most both be positive and $a >> b \geq 0$ as shown before.
This implies that both λ_{min} and λ_{max} can't be 0. If they were, then a + b = $\lambda_{max} + \lambda_{min} = 0$ which is a contradiction.
If $\lambda_{min} = 0$, we are done, $\text{cond}_2(\nabla^2 f(x^*)) = -\infty$ and so it's ill conditioned.
Else find the eigenvalues.

$$\begin{aligned} det(\nabla^2 f(x^*) - \lambda) &= 0 \\ (a-\lambda)(b-\lambda) - s^2 &= 0 \\ \lambda^2 - (a+b)\lambda + (ab - s^2) &= 0 \\ \lambda &= \frac{(a+b) \pm \sqrt{(a+b)^2 - 4(ab-s^2)}}{2} \\ &= \frac{(a+b) \pm \sqrt{a^2 + 2ab + b^2 - 4ab + 4s^2}}{2} \\ &= \frac{(a+b) \pm \sqrt{a^2 - 2ab + b^2 + 4s^2}}{2} \\ &= \frac{(a+b) \pm \sqrt{(a-b)^2 + 4s^2}}{2} \end{aligned}$$

$$\frac{\lambda_{max}}{\lambda_{min}} = \frac{(a+b) + \sqrt{(a-b)^2 + 4s^2}}{(a+b) - \sqrt{(a-b)^2 + 4s^2}}$$

But $\sqrt{(a-b)^2 + 4s^2}$ is a large number

$$\begin{aligned} a &>> b \\ a - b &>> 0 \\ (a-b)^2 &>>>> 0 \\ (a-b)^2 + 4s^2 &>>>> 0 \\ \sqrt{(a-b)^2 + 4s^2} &>> 0 \end{aligned}$$

If $s = 0$ then $\frac{\lambda_{max}}{\lambda_{min}} = \frac{a+b+(a-b)}{a+b-(a-b)} = \frac{a}{b} >> 1$, if $s \neq 0$, then the fraction will be even bigger.
Hence $\nabla^2 f(x^*)$ is ill-conditioned (because $\text{cond}_2(A) = \frac{\lambda_{max}}{\lambda_{min}}$).

6 Question 6

Do question 2.16 on page 29 of your textbook
Consider the sequence $\{x_k\}$ defined by

$$x_k = \begin{cases} (1/4)^{2^k} & \text{k even,} \\ (x_{k-1})/k & \text{k odd.} \end{cases}$$

Is this sequence Q-superlinearly convergent? Q-quadratically convergent? R-quadratically convergent?

First clearly, $\{x_k\} \to 0$ and all x_k are positive
Q-superlinearly convergent means that

$$\lim_{k\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = 0$$

For k even

$$\lim_{k\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \lim_{k\to\infty} \frac{(x_{k+1-1})/(k+1)}{x_k} = \lim_{k\to\infty} \frac{1}{k+1} = 0$$

For k odd

$$\lim_{k\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \lim_{k\to\infty} \frac{x_{k+1}}{(x_{k-1})/k} = \lim_{k\to\infty} \frac{(1/4)^{2^{k+1}}}{(1/4)^{2^{k-1}}/k} = \lim_{k\to\infty} k\frac{(1/4)^{2^{k-1}2^2}}{(1/4)^{2^{k-1}}} = \lim_{k\to\infty} k\frac{[(1/4)^{2^{k-1}}]^{2^2}}{(1/4)^{2^{k-1}}} = \lim_{k\to\infty} k[(1/4)^{2^{k-1}}]^3 = 0$$

Since $\lim_{k\to\infty} k/[4^{3*2^{k-1}}] = 0$ (∞/∞, apply l'hospital rule and get $1/\infty$)
So the sequence is Q-superlinearly convergent.

-

Q-quadratically convergent means that

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} \le M, \qquad \text{for all k sufficiently large}$$

For k even

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \frac{(x_k)/(k+1)}{x_k^2} = \frac{1}{(k+1)x_k} = \frac{1}{(k+1)(1/4)^{2^k}} = \frac{4^{2^k}}{(k+1)}$$

Which is not bounded, for large k, 4^{2^k} is much larger than $k+1$
So the sequence is not Q-quadratically convergent.

-

R-quadratically convergent means there is a sequence $\{v_k\}$ such that

$$|x_k - x^*| \le v_k \text{ for all } k \text{ and } \{v_k\} \text{ converges Q-quadratically to zero}$$

For k even

$$|x_k - x^*| = x_k = \left(\frac{1}{4}\right)^{2^k}$$

For k odd

$$|x_k - x^*| = (x_{k-1})/k = \frac{1}{k}\left(\frac{1}{4}\right)^{2^{k-1}} = \frac{1}{k}\left(\frac{1}{4}\right)^{2^{-1}2^k} = \frac{1}{k}\left(\left(\frac{1}{4}\right)^{2^{-1}}\right)^{2^k} = \frac{1}{k}\left(\frac{1}{2}\right)^{2^k}$$

Choose the sequence $v_k = \left(\frac{1}{2}\right)^{2^k}$, which clearly converges to 0 and satisfies $|x_k - x^*| \le v_k$.

$$\left(\frac{1}{4}\right)^{2^k} \le \left(\frac{1}{2}\right)^{2^k} \qquad \text{and} \qquad \frac{1}{k}\left(\frac{1}{2}\right)^{2^k} \le \left(\frac{1}{2}\right)^{2^k}$$

Must show

$$\frac{|v_{k+1} - v^*|}{|v_k - v^*|^2} \le M, \qquad \text{for all k sufficiently large}$$

$$\frac{|v_{k+1} - v^*|}{|v_k - v^*|^2} = \frac{(1/4)^{2^{k+1}}}{[(1/4)^{2^k}]^2} = \frac{(1/4)^{2^{k+1}}}{(1/4)^{2^k 2}} = \frac{(1/4)^{2^{k+1}}}{(1/4)^{2^{k+1}}} = 1 \qquad \text{clearly } M = 1 \text{ works}$$

So $\{v_k\}$ does converge Q-quadratically to zero.
Hence the sequence is R-quadratically convergent.