1 Question 1

Assume that $f : D \to \mathbf{R}$ is continuously differentiable for all $x \in D$, where the domain D of f is an open, convex subset of $\mathbf{R}^n$.
Show that f is strictly convex on D if and only if $f(y) > f(x) + \nabla f(x)^T(y-x)$ for all x and $y \in D$ for which $x \neq y$

"$\Rightarrow$" Assume f is strictly convex on D, $\forall x, y \in D$, $x \neq y$ and $\forall \alpha \in (0,1)$, have $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$
WTP $f(y) > f(x) + \nabla f(x)^T(y-x)$ for all $x, y \in D$ for which $x \neq y$

Assume for a contradiction that $\exists x, y \in D$, $x \neq y$ and $f(y) \leq f(x) + \nabla f(x)^T(y-x)$
Since f is strictly convex, f is also convex. Therefore by theorem 1, $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ $\forall x, y \in D$
So it must be that $f(y) = f(x) + \nabla f(x)^T(y-x)$ (assumed that $f(y) \leq f(x) + \nabla f(x)^T(y-x)$)
Since $\nabla f(x)$ exists and is continuous, by Taylor's Theorem

$$f(x + \alpha(y-x)) = f(x) + \alpha \nabla f(x + t\alpha(y-x))^T(y-x) \qquad \text{for some } t \in (0,1)$$

Also

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

$$\begin{aligned} f(x) + \alpha \nabla f(x + t\alpha(y-x))^T(y-x) &< \alpha f(y) + (1-\alpha)f(x) \\ \alpha \nabla f(x + t\alpha(y-x))^T(y-x) &< \alpha f(y) - \alpha f(x) \\ \nabla f(x + t\alpha(y-x))^T(y-x) &< f(y) - f(x) \end{aligned}$$

That is

$$f(y) > f(x) + \nabla f(x + t\alpha(y-x))^T(y-x) \qquad \forall \alpha \in (0,1), t \text{ is a function of } \alpha$$
$$\lim_{\alpha \to 0} \nabla f(x + t\alpha(y-x))^T(y-x) = \nabla f(x)^T(y-x) \qquad \text{since } t\alpha < \alpha \text{ and } \nabla f \text{ is continuous}$$

However this isn't enough to conclude that at the limit the $>$ inequality holds, as pointed out in the assignment handout. The assumption we want to contradict might still hold.
Instead consider the fact from theorem 1, $f(s) \geq f(t) + \nabla f(t)^T(s-t)$ $\forall t, s \in D$
This holds for

$$f(\frac{1}{2}(x+y)) \geq f(x) + \nabla f(x)^T(\frac{1}{2}(x+y) - x) = f(x) + \nabla f(x)^T(\frac{1}{2}(y-x)) = f(x) + \frac{1}{2}\nabla f(x)^T((y-x))$$

But again using the fact that f is strictly convex. And the assumed fact $f(y) = f(x) + \nabla f(x)^T(y-x)$

$$f(\frac{1}{2}(x+y)) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = \frac{1}{2}\left(f(x) + f(y)\right) = \frac{1}{2}\left(f(x) + f(x) + \nabla f(x)^T(y-x)\right) = f(x) + \frac{1}{2}\nabla f(x)^T(y-x)$$

So 'something' $\leq f(\frac{1}{2}(x+y)) <$ 'something'
Which is a contradiction Therefore the original assumption that $\exists x, y \in D$ s.t. $x \neq y$ and $f(y) \leq f(x) + \nabla f(x)^T(y-x)$ is false.
So $f(y) > f(x) + \nabla f(x)^T(y-x)$ for all $x, y \in D$ for which $x \neq y$

"$\Leftarrow$" Assume $f(y) > f(x) + \nabla f(x)^T(y-x)$ for all x and $y \in D$ for which $x \neq y$
WTP that f is strictly convex on D, $\forall x, y \in D$, $x \neq y$ and $\forall \alpha \in (0,1)$, have $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$

Let $x, y \in D$ and $x \neq y$, and let $\alpha \in (0,1)$
Let $z = \alpha x + (1-\alpha)y$, $z \in D$ since $x, y \in D$ and D is an open convex set.
If $z = x$, then $\alpha x + (1-\alpha)y = x \implies (1-\alpha)y = x - \alpha x \implies (1-\alpha)y = (1-\alpha)x \implies y = x$ since $\alpha < 1, 1-\alpha \neq 1$
If $z = y$, then $\alpha x + (1-\alpha)y = y \implies \alpha x = y - y + \alpha y \implies \alpha x = \alpha y \implies x = y$ since $\alpha > 1$, $\alpha \neq 0$
So it must be that $z \neq x$ and $z \neq y$, therefore by assumption

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

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

$(1-\alpha)f(y) > (1-\alpha)f(z) - \alpha(1-\alpha)\nabla f(z)^T(x-y)$ $\qquad$ $\alpha f(x) > \alpha f(z) + \alpha(1-\alpha)\nabla f(z)^T(x-y)$

Now adding the two equations together

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

So since all $x \neq y$ and $\alpha \in (0,1)$ were all arbitrary.
$\forall x, y \in D$, $x \neq y$ and $\forall \alpha \in (0,1)$, have $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$
That is f is a strictly convex function on D.

2 Question 2

The BFGS update

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

where

$$s_k = x_{k+1} - x_k$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Prove that if B_0 is symmetric positive-definite and $y_k^T s_k > 0$ for all $k = 0, 1, 2, \ldots$, then B_k is symmetric positive-definite for all $k = 0, 1, 2, \ldots$

First will prove this fact
If Q is symmetric positive-definite and $x \neq \alpha y$ for all $\alpha \in \mathbf{R}$ then $(x^T Q y)^2 < (x^T Q x)(y^T Q y)$ (1)
<u>Proof:</u>
We have seen in class that a symmetric positive-definite (SPD) matrix Q, has factorization of the form $Q = UDU^T$ where U is an orthogonal matrix and D is a diagonal matrix of the eigenvalues of Q, ie $U^T U = UU^T I$ and $D = \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$ where $\lambda_i > 0$ for all i because Q is SPD.
Can define matrix $D^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots \sqrt{\lambda_n})$ and also $Q^{1/2} = UD^{1/2}U^T$ which is SPD.
Clearly $D^{1/2}D^{1/2} = D$, diagonal entries multiply $\sqrt{\lambda_i}\sqrt{\lambda_i} = \lambda_i$
And $Q^{1/2}Q^{1/2} = UD^{1/2}U^T UD^{1/2}U^T = UD^{1/2}D^{1/2}U^T = UDU^T = Q$

If $x \neq \alpha y$ for all $\alpha \in \mathbf{R}$, then $Q^{1/2}x \neq \alpha Q^{1/2}y$ for all $\alpha \in \mathbf{R}$
If for some $\alpha \in \mathbf{R}$, $Q^{1/2}x = \alpha Q^{1/2}y$, then $Q^{-1/2}Q^{1/2}x = \alpha Q^{-1/2}Q^{1/2}y \implies x = \alpha y$ which is a contradiction. (since $Q^{1/2}$ is SPD, it is also invertible)
Will use Cauchy inequality below, $\left|u^T v\right| \leq \|u\|_2 \|v\|_2$ and $\left|u^T v\right| = \|u\|_2\|v\|_2 \iff u = \alpha v$

$$\begin{aligned}
(x^T Q y)^2 &= (x^T Q^{1/2} Q^{1/2} y)^2 && Q^{1/2}Q^{1/2} = Q \\
&= \left((Q^{1/2}x)^T(Q^{1/2}y)\right)^2 && Q^{1/2} \text{ is symmetric} \\
&< \left\|Q^{1/2}x\right\|_2^2 \left\|Q^{1/2}y\right\|_2^2 && \text{by Cauchy inequality, } u = Q^{1/2}x \text{ and } v = Q^{1/2}y,\ u \neq \alpha v \\
&= \left((Q^{1/2}x)^T(Q^{1/2}x)\right)\left((Q^{1/2}y)^T(Q^{1/2}y)\right) && \text{by def of 2 norm} \\
&= \left(x^T Q^{1/2}Q^{1/2}x\right)\left(y^T Q^{1/2}Q^{1/2}y\right) && Q^{1/2} \text{ is symmetric} \\
&= \left(x^T Q x\right)\left(y^T Q y\right) && Q^{1/2}Q^{1/2} = Q
\end{aligned}$$

Now, define $P(k)$: B_k as defined by the above BFGS update is symmetric positive-definite
<u>Base Case:</u> $P(0)$ holds by given assumption, B_0 is symmetric positive-definite
<u>Induction Step:</u> Assume $P(k)$ holds, that is B_k is SPD. WTP $P(k)$ holds, that B_{k+1} is SPD
By def of SPD, need to show that for all $x \neq 0 \in \mathbf{R}^n$, $x^T B_{k+1} x > 0$

Let $x \in \mathbf{R}^n$ and $x \neq 0$

$$
\begin{aligned}
x^T B_{k+1} x &= x^T B_k x - x^T \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} x + x^T \frac{y_k y_k^T}{y_k^T s_k} x && \text{by def of } B_{k+1} \\
&= x^T B_k x - \frac{x^T B_k s_k s_k^T B_k x}{s_k^T B_k s_k} + \frac{x^T y_k y_k^T x}{y_k^T s_k} \\
&= x^T B_k x - \frac{(s_k^T B_k x)^T (s_k^T B_k x)}{s_k^T B_k s_k} + \frac{(y_k^T x)^T (y_k^T x)}{y_k^T s_k} && \text{by prop of } T \text{ and matrix mult associativity} \\
&= x^T B_k x - \frac{(s_k^T B_k x)^2}{s_k^T B_k s_k} + \frac{(y_k^T x)^2}{y_k^T s_k} && \text{the values are scalars, } a^T = a \text{ when } a \in \mathbf{R}
\end{aligned}
$$

Case 1: If $x = \alpha s_k$ for some $\alpha \in \mathbf{R}$, clearly $\alpha \neq 0$, otherwise x would be 0
Then

$$
\begin{aligned}
x^T B_{k+1} x &= (\alpha s_k)^T B_k (\alpha s_k) - \frac{(s_k^T B_k (\alpha s_k))^2}{s_k^T B_k s_k} + \frac{(y_k^T (\alpha s_k))^2}{y_k^T s_k} \\
&= \alpha^2 s_k^T B_k s_k - \frac{\alpha^2 (s_k^T B_k s_k)^2}{s_k^T B_k s_k} + \frac{\alpha^2 (y_k^T s_k)^2}{y_k^T s_k} \\
&= \alpha^2 s_k^T B_k s_k - \alpha^2 (s_k^T B_k s_k) + \alpha^2 (y_k^T s_k) \\
&= \alpha^2 (y_k^T s_k) \\
&> 0 && y_k^T s_k > 0 \text{ by assumption, and } \alpha^2 > 0 \text{ since } \alpha \neq 0
\end{aligned}
$$

Case 2: If $x \neq \alpha s_k$ for all $\alpha \in \mathbf{R}$
Then

$$
\begin{aligned}
x^T B_{k+1} x &= x^T B_k x - \frac{(s_k^T B_k x)^2}{s_k^T B_k s_k} + \frac{(y_k^T x)^2}{y_k^T s_k} \\
&> x^T B_k x - \frac{(x^T B_k x)(s_k^T B_k s_k)}{s_k^T B_k s_k} + \frac{(y_k^T x)^2}{y_k^T s_k} && \text{by I.H. } B_k \text{ is SPD and so line follows from prop } (\mathbf{1}) \\
&= x^T B_k x - (x^T B_k x) + \frac{(y_k^T x)^2}{y_k^T s_k} \\
&= \frac{(y_k^T x)^2}{y_k^T s_k} \\
&\geq 0 && \text{since } y_k^T s_k > 0 \text{ and } (y_k^T x)^2 \geq 0
\end{aligned}
$$

In either case, $x^T B_{k+1} x > 0$ and since x was arbitrary, it is true for all $x \neq 0$.
Therefore B_{k+1} is SPD and so $P(k+1)$ holds.

This proves that $P(k)$ holds for all k, that is B_k is SPD for all $k = 0, 1, 2...$ (as long as for all $k = 0, 1, 2..$, $y_k^T s_k > 0$)

3 Question 3

Show that if all the following is true
a) $\nabla f(x_k) \neq 0$ (i.e., you are not already at a critical point)
b) $\nabla^2 f(x)$ exists, is continuous and is symmetric positive-definite for all $x \in D$, where D is an open, convex set containing both x_k and x_{k+1}.
c) B_k is symmetric positive-definite,
d) $x_{k+1} = x_k + p_k$, where $p_k = -B_k^{-1} \nabla f(x_k)$
Then $y_k^T s_k > 0$, where y_k and s_k are given in Question 2 above.

By b) the convex set D contains x_k and $x_{k+1} = x_k + p_k$, so it must contain $\alpha x_{k+1} + (1 - \alpha) x_k$ for all $\alpha \in [0, 1]$
$\alpha x_{k+1} + (1 - \alpha) x_k = \alpha (x_k + p_k) + (1 - \alpha) x_k = \alpha x_k + \alpha p_k + x_k - \alpha x_k = x_k + \alpha p_k \in D$
That is $\nabla^2 f(x_k + \alpha p_k)$ exists, is continuous and is symmetric positive-definite for all $\alpha \in [0, 1]$ **(2)**

By Taylor Theorem seen in class, whose continuity conditions on $\nabla^2 f(x)$ hold by above argument, we can write
$\nabla f(x_{k+1}) = \nabla f(x_k + p_k) = \nabla f(x_k) + \int_{t=0}^{1} \nabla^2 f(x_k + tp_k) p_k dt$
So $\nabla f(x_{k+1}) - \nabla f(x_k) = \int_{t=0}^{1} \nabla^2 f(x_k + tp_k) p_k dt$

$$
\begin{aligned}
y_k^T s_k = s_k^T y_k &= (x_{k+1} - x_k)^T \left[\nabla f(x_{k+1}) - \nabla f(x_k)\right] \\
&= p_k^T \left[\int_{t=0}^{1} \nabla^2 f(x_k + tp_k) p_k dt\right] && \text{by above } \textbf{(2)} \text{ and d) } x_k = x_k + p_k \\
&= \int_{t=0}^{1} p_k^T \nabla^2 f(x_k + tp_k) p_k dt \\
&= \int_{t=0}^{1} \left(-B_k^{-1} \nabla f(x_k)\right)^T \nabla^2 f(x_k + tp_k) \left(-B_k^{-1} \nabla f(x_k)\right) dt && \text{by d)} \\
&= \int_{t=0}^{1} \left(B_k^{-1} \nabla f(x_k)\right)^T \nabla^2 f(x_k + tp_k) \left(B_k^{-1} \nabla f(x_k)\right) dt \\
&> 0
\end{aligned}
$$

The last inequality comes from the fact that $\nabla f(x_k) \neq 0$ and that B_k is SPD, specifically it is invertible, so $B_k^{-1} \nabla f(x_k) \neq 0$ ($B_k a = 0 \implies a = 0$).
Next by **(2)**, $\nabla^2 f(x_k + tp_k)$ is SPD for all $t \in [0, 1]$, exactly the integration bounds.
Meaning the integrand

$$
\left(B_k^{-1} \nabla f(x_k)\right)^T \nabla^2 f(x_k + tp_k) \left(B_k^{-1} \nabla f(x_k)\right) > 0 \qquad \text{for the integral bounds}
$$

Integral of continuous positive function is positive, $\int_D f(x) dx > 0$ if $f(x) > 0$ for all $x \in D$