1. Since B_k is p.d, we can have $B_k = V \Delta V^T > 0$

Since $P_k = \arg\min \{ g_k^T p + \frac{1}{2} p^T B_k p \} = -B_k^{-1} g_k$, so $g_k^T P_k = -g_k^T B_k^{-1} g_k$

in $g_k^T P_k = -g_k^T B_k^{-1} g_k$, replace B_k^{-1} with $V \Delta^{-1} V^T$

we can get $g_k^T P_k = -\underbrace{g_k^T (V \Delta^{-\frac{1}{2}} V^T)}_{C^T} (V \Delta^{-\frac{1}{2}} V^T) g_k$

$= -C^T C$

$C^T C \geq \frac{1}{\lambda_{max}} \|g_k\|^2$, so $-C^T C \leq -\frac{1}{\lambda_{max}} \|g_k\|^2 \leq -\frac{1}{\lambda_{max}} \|\nabla f(x_k)\|^2$

$\|P_k\|_2^2 = \|-B^{-1} g_k\|^2 \leq \|B^{-1}\|^2 \cdot \|g_k\|^2 \leq \frac{1}{\lambda_{min}^2} \cdot \|g_k\|^2 \leq \frac{1}{\lambda_{min}^2} \|\nabla f(x_k)\|^2$

Since f is with L-Lip, so $f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2} \|y-x\|^2$, for $\forall x, y$

So $f(x_{k+1}) \leq f(x_k) + g_k^T (a P_k) + \frac{L a^2}{2} \|P_k\|^2$

$= f(x_k) + a g_k^T P_k + \frac{L a^2}{2} \|P_k\|^2$

So $f(x_{k+1}) \leq f(x_k) + \alpha g_k^T P_k + \frac{L}{2} \alpha^2 \|P_k\|_2^2$

$\leq f(x_k) + \alpha \frac{-1}{\lambda_{max}} \|g_k^2\| + \frac{L \alpha^2}{2} \cdot \frac{1}{\lambda_{min}} \|g_k\|^2$

$= f(x_k) - (\frac{\alpha}{\lambda_{max}} - \frac{L \alpha^2}{2 \lambda_{min}}) \|g_k\|^2$

b) $\frac{1}{\lambda_{max}} a - \frac{L}{2 \lambda_{min}^2} a^2 = -\frac{L}{2 \lambda_{min}^2} (a - \frac{\lambda_{min}^2}{L \cdot \lambda_{max}})^2 + \frac{\lambda_{min}^2}{2 L \cdot \lambda_{max}^2}$

so when $a = \frac{\lambda_{min}^2}{L \cdot \lambda_{max}}$ maximizes this amount, the value is positive

the maximum quantity is $\frac{\lambda_{min}^2}{2 L \cdot \lambda_{max}^2} \|g_k\|^2$ which is non-negative since $\|g_k\|^2 \geq 0$

c) from a) and b) we can have:

$$f(x_{k+1}) \le f(x_k) - \frac{\lambda_{min}^2}{2L\lambda_{max}^2} \|g_k\|^2$$

so $f(x_k) \le f(x_{k-1}) - \frac{\lambda_{min}^2}{2L\lambda_{max}^2} \|g_{k-1}\|^2$

$\vdots$

then, $f(x_{k+1}) \le f(x_0) - \frac{\lambda_{min}^2}{2L\lambda_{max}^2} \sum_{i=0}^{k} \|g_0\|^2$

$$\le f(x_0) - \frac{\lambda_{min}^2}{2L\lambda_{max}^2} \cdot (k+1) \cdot \min_{s \le k} \|g_s\|^2$$

Since $g_k = \nabla f(x_k)$,

$$\min_{i \le k} \|f(x_i)\|^2 \le \frac{2L\lambda_{max}^2}{\lambda_{min}^2 (k+1)} \cdot \left(f(x_0) - f(x_{k+1})\right)$$

$$\le \frac{2L\lambda_{max}^2}{\lambda_{min}^2 (k+1)} \cdot \left(f(x_0) - f_{min}\right)$$

so

$$\min_{i \le k} \|f(x_i)\| \le \frac{M}{\sqrt{k+1}}, \quad M = \frac{\lambda_{max}}{\lambda_{min}} \sqrt{(f(x_0) - f_{min}) \cdot 2L}$$

d) when $B_k = L \cdot I$, then $\lambda_{min} = \lambda_{max} = L$

to maximize the quadratic term: $\alpha = \frac{\lambda_{min}^2}{L\lambda_{max}} = 1$

So $$\min_{i \le k} \|\nabla f(x)\| \le \frac{\sqrt{2L(f(x_0) - f_{min})}}{\sqrt{k+1}}$$

it converges linearly

2. $B_{k+1} = B_k + \dfrac{y_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}} - \dfrac{B_k s_{k+1} (B_k s_{k+1})^T}{s_{k+1}^T B_k s_{k+1}}$

$$x^T B_{k+1} x = x^T \left(B_k + \frac{y_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}} - \frac{B_k s_{k+1} s_{k+1}^T B_k^T}{s_{k+1}^T B_k s_{k+1}} \right) x$$

$$= x^T B_k x + x^T \frac{y_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}} x - x^T \frac{B_k s_{k+1} s_{k+1}^T B_k^T}{s_{k+1}^T B_k s_{k+1}} x$$

$V^T V = B_k$ since p.d. symmetric.

$$x^T \frac{B_k s_{k+1} s_{k+1}^T B_k^T}{s_{k+1}^T B_k s_{k+1}} x = \frac{x^T V^T V s_{k+1} s_{k+1}^T V^T V x}{s_{k+1}^T V^T V s_{k+1}}$$

$$= \frac{\| x^T V^T V s_{k+1} \|^2}{\| V s_{k+1} \|^2} \leq \frac{\| x^T V^T \|^2 \cdot \| V \cdot s_{k+1} \|^2}{\| V \cdot s_{k+1} \|^2} \leq \| V x \|^2$$

$$x^T B_k x = \| V x \|^2$$

so $x^T B_k x - x^T \dfrac{B_k s_{k+1} s_{k+1}^T B_k^T}{s_{k+1}^T B_k s_{k+1}} x \geq 0$

and $x^T \dfrac{y_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}} x \geq 0$ since $y_{k+1}^T s_{k+1} > 0$ $\quad x^T y_{k+1} y_{k+1}^T x = \| x^T y_{k+1} \|^2 > 0$

so $x^T B_{k+1} x > 0$ so B_{k+1} is p.d

b) Woodbury matrix identity

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$

Since $B_{k+1} - B_k = U \leftarrow$ rank two, symmetric

So we can have $B_{k+1} = B_k + \beta vv^T + \delta ww^T$ $\beta, \delta \neq 0$, v, w linearly independent

So let $\begin{cases} A = B_k \\ U = [v, w] \in R^{d\times 2} \\ V = U^T \\ C = \begin{bmatrix} \beta & 0 \\ 0 & \delta \end{bmatrix} \end{cases}$, and we can have: $\begin{cases} C^{-1} = \begin{bmatrix} \beta^{-1} & 0 \\ 0 & \delta^{-1} \end{bmatrix} \\ C^{-1} + VA^{-1}U = \begin{bmatrix} v^TA^{-1}v + \beta^{-1}, & v^TA^{-1}w \\ w^TA^{-1}v, & w^TA^{-1}w + \delta^{-1} \end{bmatrix} \end{cases}$

Sinc $B_{k+1} - B_k = \beta vv^T + \delta ww^T = UCV$

So $A + UCV = B_{k+1}$

So we can have:

$$(B_{k+1})^{-1} = B_k^{-1} - B_k^{-1}[v,w]\begin{bmatrix} v^TB_k^{-1}v + \beta^{-1}, & v^TB_k^{-1}w \\ w^TB_k^{-1}v, & w^TB_k^{-1}w + \delta^{-1} \end{bmatrix}^{-1}\begin{bmatrix} v^T \\ w^T \end{bmatrix} B_k^{-1}$$

$$\begin{bmatrix} v^TB_k^{-1}v + \beta^{-1}, & v^TB_k^{-1}w \\ w^TB_k^{-1}v, & w^TB_k^{-1}w + \delta^{-1} \end{bmatrix}^{-1} = \frac{1}{\underbrace{(v^TB_k^{-1}v + \beta^{-1})(w^TB_k^{-1}w + \delta^{-1}) - (v^TB_k^{-1}w)^2}_{K}}\begin{bmatrix} w^TB_k^{-1}w + \delta^{-1}, & -v^TB_k^{-1}w \\ -w^TB_k^{-1}v, & v^TB_k^{-1}v + \beta^{-1} \end{bmatrix}$$

$$\text{so } [v,w]\begin{bmatrix} v^TB_k^{-1}v + \beta^{-1}, & v^TB_k^{-1}w \\ w^TB_k^{-1}v, & w^TB_k^{-1}w + \delta^{-1} \end{bmatrix}^{-1}\begin{bmatrix} v^T \\ w^T \end{bmatrix}$$

$$= \frac{1}{K}\left[\cancel{(w^TB_k^{-1}w + \delta^{-1})}\cdot vv^T - (v^TB_k^{-1}w)vw^T - (w^TB_k^{-1}v)wv^T + (v^TB_k^{-1}v + \beta^{-1})ww^T\right]$$

in BFGS $\delta = \frac{-1}{s_{k+1}^TB_ks_{k+1}}$ $\beta = \frac{1}{y_{k+1}^Ts_{k+1}}$

so let $\begin{cases} V = y_{k+1} \\ W = B_k s_{k+1} \end{cases}$, so $W^T B_k^{-1} W = s_{k+1}^T B_k s_{k+1} = -\delta^{-1}$

then we have $k = (V^T B_k^{-1} V + \beta^{-1})(W^T B_k^{-1} W + \delta^{-1}) - (V^T B_k^{-1} W)^2$

$$= -(y_{k+1}^T s_{k+1})^2$$

so $$[V, W] \begin{bmatrix} V^T B_k^{-1} V + \beta^{-1}, & V^T B_k^{-1} W \\ W^T B_k^{-1} V, & W^T B_k^{-1} W + \delta^{-1} \end{bmatrix}^{-1} \begin{bmatrix} V^T \\ W^T \end{bmatrix}$$

$$= \frac{-1}{(y_{k+1}^T s_{k+1})^2} \Big\{ - y_{k+1}^T s_{k+1} \, y_{k+1} s_{k+1}^T B_k - s_{k+1}^T y_{k+1} B_k s_{k+1} \cdot y_{k+1}^T + y_{k+1}^T B_k^{-1} y_{k+1} B_k s_{k+1} s_{k+1}^T B_k$$
$$+ y_{k+1}^T s_{k+1} B_k s_{k+1} s_{k+1}^T B_k \Big]$$

$$- B_k^{-1} [V, W] \begin{bmatrix} V^T B_k^{-1} V + \beta^{-1}, & V^T B_k^{-1} W \\ W^T B_k^{-1} V, & W^T B_k^{-1} W + \delta^{-1} \end{bmatrix}^{-1} \begin{bmatrix} V^T \\ W^T \end{bmatrix} B_k^{-1}$$

$$= \frac{1}{(y_{k+1}^T s_{k+1})^2} (y_{k+1}^T B_k^{-1} y_{k+1} + y_k^T s_k) s_k s_k^T$$

$$- \frac{1}{y_{k+1}^T s_{k+1}} (B_k^{-1} y_{k+1} s_{k+1}^T + s_{k+1} y_{k+1}^T B_k^{-1})$$

a scalar, can insert to the middle

So $$B_{k+1}^{-1} = \left(B_k^{-1} - B_k^{-1} \frac{y_{k+1} s_{k+1}^T}{y_{k+1}^T s_{k+1}} - \frac{s_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}} B_k^{-1} + \frac{s\, y_{k+1}^T B_k^{-1} y_{k+1} s_k^T}{(y_{k+1}^T s_{k+1})^2} \right)$$

$$+ \frac{s_k s_k^T}{y_{k+1}^T s_{k+1}}$$

$$= \left(I - \frac{s_{k+1} y_{k+1}^T}{y_{k+1}^T s_{k+1}}\right) B_k^{-1} \left(I - \frac{y_{k+1} s_{k+1}^T}{y_{k+1}^T s_{k+1}}\right) + \frac{s_k s_k^T}{y_{k+1}^T s_{k+1}}$$

which is DFP update