(1) (a) Supongamos que los vectores $p_0, \ldots, p_l$ son LD, esto implicaría que alguna combinación lineal de ellos equivale a cero y por ende son linealmente dependientes:

$$p_i = \lambda_0 p_0 + \ldots + \lambda_l p_l$$

para alguna cantidad de coeficientes λ_k
$k \in \{0, \ldots, l\} / i$ (se excluye a i).

Entonces

$$0 = p_0^T A p_i = \lambda_0 p_0^T A p_0 + \ldots + \lambda_l p_0^T A p_l$$

$$= \lambda_0 p_0^T A p_0.$$

Como $p_i^T A p_i \neq 0$ porque A es positiva definida, entonces $\lambda_0 = 0$. Análogamente $\lambda_k = 0 \ \forall k$.
Esto implica que

$$p_i = 0 + \ldots + 0 = 0$$

lo cual contradice el hecho de que los vectores son no nulos. Por ende los vectores son LI.

(b) Dado que los vectores son LI $\mathrm{span}\{x_1, \ldots, x_n\}$ abarca $\mathbb{R}^n$.

Podemos escribir

$$x^* - x_0 = \sigma_0 p_0 + \ldots + \sigma_{n-1} p_{n-1}$$

Entonces tenemos que

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k}$$

los cuales coinciden con

$$\alpha_k = \frac{-r_k^T p_k}{p_k^T A p_k}$$

pues

$$x_k = x_0 + \alpha_0 p_0 + \ldots + \alpha_{k-1} p_{k-1}$$

Al multiplicar la expresión por $p_k^T A$:

$$p_k^T A(x_k - x_0) = 0$$

$$\Rightarrow p_k^T A(x^* - x_0) = p_k^T A(x^* - x_k) = p_k^T(b - Ax_k)$$

$$= -p_k^T r_k.$$

Con esto queda demostrado que $\alpha_k = \sigma_k$ y por eso el algoritmo converge en a lo más n iteraciones.

② ⓐ La 2da condición fuerte de Wolfe es

$$\left|\nabla f(x_k + \alpha_k p_k)^T p_k\right| \leq c_2 \left|\nabla f(x_k)^T p_k\right|$$

$$\Rightarrow \nabla f(x_k + \alpha_k p_k)^T p_k \geq -c_2 \left|\nabla f(x_k)^T p_k\right|$$

$$= c_2 \nabla f(x_k)^T p_k$$

(Como p_k es dirección de descenso $\left|\nabla f(x_k)^T p_k\right| = -\nabla f(x_k)^T p_k$).

$$\Rightarrow \nabla f(x_k + \alpha_k p_k)^T p_k - \nabla f(x_k)^T p_k$$

$$= (c_2 - 1) \nabla f(x_k)^T p_k > 0 \quad \Big| \cdot \alpha_k ; \ (c_2 < 1)$$

$$\Rightarrow \nabla f(x_k + \alpha_k p_k)^T p_k \alpha_k - \nabla f(x_k)^T p_k \alpha_k$$

$$= (c_2 - 1) \nabla f(x_k)^T p_k \alpha_k > 0 \qquad (p_k \alpha_k = s_k)$$

$$\Rightarrow \nabla f(x_k + s_k)^T s_k - \nabla f(x_k)^T s_k$$

$$= (c_2 - 1) \nabla f(x_k)^T s_k > 0 \qquad (y_k = \nabla f(x_k + \alpha_k p_k) - \nabla f(x_k))$$

$$y_k^T s_k = (c_2 - 1) f(x_k)^T s_k > 0$$

$$\Rightarrow s_k^T y_k > 0 \quad \blacksquare$$

b

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

(b) $B_{k+1} H_{k+1} = B_{k+1} \left((I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) \right.$

$\left. + \rho_k s_k s_k^T \right)$

$= (B_{k+1} - \rho_k y_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$

$\overset{*}{=} \left(B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$

$B_k H_k = I$

$= \left(I - \dfrac{B_k s_k s_k^T}{s_k^T B s_k} \right) (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$

$= I - \dfrac{B_k s_k s_k^T}{s_k^T B_k s_k} - \rho_k y_k s_k^T + \dfrac{B_k s_k s_k^T}{s_k^T B_k s_k} + \rho_k y_k s_k^T$

$= I$

(*) En este paso se acortó esto:

$\rho_k \dfrac{B_k s_k s_k^T}{s_k^T B_k s_k} y_k s_k^T = \dfrac{1}{s^T y_k} \dfrac{B_k s_k (s_k^T y_k) s_k^T}{s_k^T B_k s_k}$

$= \dfrac{B_k s_k s_k^T}{s_k^T B_k s_k}$