1. (a) $\alpha^k = \underset{\alpha}{\arg\min}\, f(x^k + \alpha d^k)$ with $d^k = -\nabla f(x^k)$

Since α^k minimizes $h(\alpha) = f(x^k + \alpha d^k)$, $h'(\alpha)$ must be 0 at $\alpha = \alpha^k$

Then, we have $h'(\alpha) = \dfrac{d f(x^k + \alpha d^k)}{d\alpha} = \langle d^k, \nabla f(x^k + \alpha d^k) \rangle$

Thus, $h'(\alpha^k) = \langle d^k, \nabla f(x^k + \alpha^k d^k) \rangle = 0$

(b) Since $d^k = -\nabla f(x^k)$,

$$\langle d^k, \nabla f(x^k + \alpha^k d^k) \rangle = \langle -\nabla f(x^k), \nabla f(x^k + \alpha^k d^k) \rangle = 0$$

x_k and x_{k+1} are two consecutive points generated by the steepest descent algorithm, then we have

$$x^{k+1} = x^k + \alpha^k d^k$$

Thus, $\langle -\nabla f(x^k), \nabla f(x^{k+1}) \rangle = 0$

$$\langle \nabla f(x^k), \nabla f(x^{k+1}) \rangle = 0$$

3. (a) $f(x) = \frac{x_1^4}{4} - x_1^2 + 2x_1 + (x_2-1)^2$

$$(\nabla f)(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_1^3 - 2x_1 + 2 \\ 2(x_2-1) \end{pmatrix}$$

$$(Hf)(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 3x_1^2 - 2 & 0 \\ 0 & 2 \end{pmatrix}$$

According to Newton's method, $x^{k+1} = x^k - (Hf)(x^k)^{-1}(\nabla f)(x^k)$

For $k=0$, we have $x_0 = \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix}$

$$(Hf)(x_0) = \begin{pmatrix} 3x_{0,1}^2 - 2 & 0 \\ 0 & 2 \end{pmatrix}$$

According to Newton's method, we need $(Hf)(x^k)$ to be invertible at each iteration. $Hf(x_0)$ is invertible iff $\det((Hf)(x_0)) \neq 0$

Thus, $\det((Hf)(x_0)) = 6x_{0,1}^2 - 4 = 0$

$$\Longrightarrow x_{0,1} = \pm\sqrt{\frac{2}{3}}$$

Therefore, the initialization value $x_0 = \begin{pmatrix} \pm\sqrt{\frac{2}{3}} \\ x_{0,2} \end{pmatrix}$, $x_{0,2} \in R$ are points where $(Hf)(x_0)$ is non-invertible and the Newton's method breaks down at these points.