Exercise Sheet 4

(a) Let the gradient $\underline{g}_t := \underline{\nabla} E^T_{[\underline{w}_t]}$ and the Hesse matrix $\underline{H}_t := \underline{\Delta} E^T_{[\underline{w}_t]}$ then

$$E^T_{[\underline{w}_{t+1}]} = E^T_{[\underline{w}_t]} + (\underline{w}_{t+1} - \underline{w}_t)^T \underline{g}_t + \frac{1}{2}(\underline{w}_{t+1} - w_t)^T H_t (w_{t+1} - w_t)$$

$$= E^T_{[w_t]} - \eta_t \underline{d}_t^T g_t + \frac{\eta^2}{2} \underline{d}_t \underline{H}_t \underline{d}_t$$

(b) $E^T_{[\underline{w}_{t+1}]} \approx E^T_{[\underline{w}_t]} - \eta_t \underline{d}_t^T \underline{g}_t + \frac{\eta^2}{2} \underline{d}_t^T \underline{H}_t \underline{d}_t \overset{!}{\leq} E^T_{[\underline{w}_t]}$

$$\Longrightarrow \eta_t \overset{!}{\leq} 2 \frac{\underline{d}_t^T \underline{g}_t}{\underline{d}_t^T \underline{H}_t \underline{d}_t}$$

(c) Solve $\min_\eta E^T(\underline{w}_{t+1})$ by setting the derivative w.r.t η to zero:

$$\frac{\partial E^T_{[\underline{w}_{t+1}]}}{\partial \eta} = \left(\frac{\partial E^T_{[\underline{w}_{t+1}]}}{\partial \underline{w}_{t+1}}\right)^T \frac{\partial \underline{w}_{t+1}}{\partial \eta}$$

$$= \left(\underline{H}_t w_t - \eta_t \underline{H}_t \underline{d}_t - \underline{H}_t \underline{w}^*\right)^T (-\underline{d}_t) \overset{!}{=} 0$$

$$\Longrightarrow \eta^* = \frac{\underline{d}_t^T \underline{H}_t (\underline{w}_t - \underline{w}^*)}{\underline{d}_t^T \underline{H}_t \underline{d}_t} = \frac{\underline{d}_t^T \underline{g}_t}{\underline{d}_t^T \underline{H}_t \underline{d}_t}$$

as $g_t = \underline{H}_t (\underline{w}_t - \underline{w}^*)$

(d) The gradient is orthogonal to the direction if $\underline{d}_t^T g_{t+1} = 0$

$$g_{t+1} = H_t(\underline{w}_{t+1} - w^*) = H_t(w_t - \eta_t \underline{d}_t - w^*)$$

$$= g_t - \eta_t \underline{H}_t \underline{d}_t$$

$$\underline{d}_t^T g_{t+1} = \underline{d}_t^T g_t - \frac{\underline{d}_t^T g_t}{\underline{d}_t^T \underline{H}_t \underline{d}_t} \underline{d}_t^T \underline{H}_t \underline{d}_t = 0$$