Newton Method

The Newton method uses information about the second order derivatives of the function $f(\vec{x})$. If we assume that function $f(\vec{x})$ can be locally represented by its second order Taylor expansion, we can write:

$$\begin{aligned} f(\vec{x}(t+1)) &= f(\vec{x}(t)) \; + \; (\vec{x}(t+1) - \vec{x}(t))^T \, \nabla f(x(\vec{t})) \\ & \quad + \frac{1}{2} \, (\vec{x}(t+1) - \vec{x}(t))^T \, \nabla^2 f(x(\vec{t})) \, (\vec{x}(t+1) - \vec{x}(t)). \end{aligned} \quad (3)$$

The term $\nabla^2 f(x(\vec{t}))$ is the Hesse matrix at point $x(t)$: $\nabla^2 f(x(\vec{t})) = \mathcal{H}(f(\vec{x}(t)))$. Since the target is to reach the extremum at time step $t+1$, we have

$$\nabla f(\vec{x}(t+1)) \stackrel{!}{=} 0 \quad (4)$$

$$\Rightarrow$$

$$\nabla f(x(\vec{t})) \; + \; \mathcal{H}(f(\vec{x}(t))(\vec{x}(t+1) - \vec{x}(t)) \; = \; 0 \quad (5)$$

$$\vec{x}(t+1) = \vec{x}(t) - \mathcal{H}^{-1}(f(\vec{x}(t))\nabla f(x(\vec{t})) \quad (6)$$

Therefore, the direction $\vec{s}(t)$ is given by:

$$\vec{s}(t) \; = \; -\mathcal{H}^{-1}(f(\vec{x}(t))\nabla f(x(\vec{t})) \quad (7)$$

Indeed for any second order function, the Newton descent reaches the optimum after one iteration for any start point.