

If the gradient of f exists, it holds for every τ, q that

$$f(\mathbf{p} + \tau q) = f(\mathbf{p}) + \tau \nabla f(\mathbf{p})^T \cdot q + r(\tau)$$

with $\lim_{\tau \rightarrow 0} \frac{r(\tau)}{\tau} = 0$. Hence we get for h

$$\begin{aligned} h(\mathbf{p} + \tau q) &= g(f(\mathbf{p} + \tau q)) = g(f(\mathbf{p}) + \tau \nabla f(\mathbf{p})^T \cdot q + r(\tau)) \\ &= g(f(\mathbf{p}) + \tau (\nabla f(\mathbf{p})^T \cdot q + \frac{r(\tau)}{\tau})) \\ &= g(f(\mathbf{p})) + \tau g'(f(\mathbf{p})) \cdot (\nabla f(\mathbf{p})^T \cdot q + \frac{r(\tau)}{\tau}) + r'(\tau) \\ &= h(\mathbf{p}) + \tau g'(f(\mathbf{p})) \cdot \nabla f(\mathbf{p})^T \cdot q + \tau g'(f(\mathbf{p})) \frac{r(\tau)}{\tau} + r'(\tau) . \end{aligned}$$

It clearly is

$$\lim_{\tau \rightarrow 0} \frac{\tau g'(f(\mathbf{p})) \frac{r(\tau)}{\tau} + r'(\tau)}{\tau} = \lim_{\tau \rightarrow 0} g'(f(\mathbf{p})) \cdot \frac{r(\tau)}{\tau} + \frac{r'(\tau)}{\tau} = 0$$

and thus

$$h(\mathbf{p} + \tau q) = h(\mathbf{p}) + \tau \nabla h(\mathbf{p})^T \cdot q + r''(\tau)$$

with $r''(\tau) = \tau g'(f(\mathbf{p})) \frac{r(\tau)}{\tau} + r'(\tau)$ and $\nabla h(\mathbf{p}) = g'(f(\mathbf{p})) \cdot \nabla f(\mathbf{p})$ which was to show.