Exercise 1.3: Gradients and gradient descent

(a) Given: $E(\theta) := 2\theta_1^2 + 4\theta_2 + \max(0, \theta_2 + \theta_3)$, $\quad \theta^{[0]} = [2, 1, 0]^T$, $\quad \tau = 0.5$

$$\nabla E = (\frac{\partial E}{\partial \theta_1}, \frac{\partial E}{\partial \theta_2}, \frac{\partial E}{\partial \theta_3})^T$$

$$\frac{\partial E}{\partial \theta_1} = 4\theta_1, \quad \frac{\partial E}{\partial \theta_2} = \begin{cases} 4+1 & , \quad \theta_2 \geq 0 \\ 4+0 & , \quad \theta_2 < 0 \end{cases}, \quad \frac{\partial E}{\partial \theta_3} = \begin{cases} 1 & , \quad \theta_3 \geq 0 \\ 0 & , \quad \theta_3 < 0 \end{cases}$$

$$\begin{aligned} \theta \to \theta' &= \theta - \tau \nabla E \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 4 \cdot 2 \\ 4+1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ -1.5 \\ -0.5 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \theta' \to \theta'' &= \theta' - \tau \nabla E \\ &= \begin{bmatrix} -2 \\ -1.5 \\ -0.5 \end{bmatrix} - 0.5 \begin{bmatrix} 4 \cdot (-2) \\ 4+0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ -3.5 \\ -0.5 \end{bmatrix} \end{aligned}$$

(b) $f : \mathbb{R}^n \mapsto \mathbb{R}, \quad g : \mathbb{R} \mapsto \mathbb{R}, \quad h := g \circ f, \quad h(p) = g(f(p))$
Show that:

$$\nabla h(p) = g'(f(p)) \nabla f(p)$$

Assumption: all derivatives exist
Proof:

$$\begin{aligned} f(x + \alpha h) &= f(x) + \alpha (Df)(x)^T h + r_f(\alpha), && \text{with } \lim_{\alpha \to 0} \tfrac{r_f(\alpha)}{\alpha} = 0 \\ g(y + \beta) &= g(q) + \beta (Dg)(q) + r_g(\beta), && \text{with } \lim_{\beta \to 0} \tfrac{r_g(\beta)}{\beta} = 0 \\ h(p + \tau q) &= g(f(p + \tau q)) \\ &= g(\underbrace{f(p)}_{y} + \underbrace{\tau (Df)(p)^T q + r_f(\tau)}_{\beta}) \\ &= g(f(p)) + (Dg)(f(p)) \cdot (\tau (Df)(p)^T q + r_f(\tau)) + r_g(\tau (Df)(p)^T q + r_f(\tau)) \\ \Rightarrow \quad &= \underbrace{(g \circ f)(p)}_{h(p)} + \underbrace{(Dg)(f(p))}_{g'(f(p))} \underbrace{(\tau (Df)(p)^T q)}_{\nabla f(p)} + \underbrace{(Dg)(f(p)) r_f(\tau)}_{0 \text{ for } \tau \to 0} + \underbrace{r_g(\tau (Df)(p)^T q + r_f(\tau))}_{0 \text{ for } \tau \to 0} \end{aligned}$$