Name: Mohammad Farhad.
Reg: 2017831008

ⓐ $f(z) = \log(1+z)$

where, $z = x^T x \quad x \in R^d$.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \quad x^T = [x_1 \; x_2 \cdots x_d]$$

$$x^T x = [x_1 \; x_2 \cdots x_d] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$$= [x_1^2 + x_2^2 + \cdots + x_d]^2.$$

using chain rule,

$$\frac{df}{dx} = \frac{df}{dz} \cdot \frac{dz}{dx}$$

$$= \frac{d}{dz} \log(1+z) \cdot \frac{d}{dx}(x^T x).$$

$$= \frac{1}{1+z} \frac{d}{dx}\left([x_1^2 + x_2^2 + \cdots x_d^2]\right)$$

$$= \left(\frac{1}{1+x^T x}\right) \cdot 2 \sum_{i=1}^{d} x_i$$

So the gradient of the first equation is

$$\left(\frac{2}{1+x^T x}\right) \sum_{i=1}^{d} (x_i).$$

(b) $f(z) = e^{-z/2}$

where, $z = g(y) = y^T S^{-1} y$

$y = h(x) = x - \mu$

$x, \mu \in R^d, \; S \in R^{d \times d}$.

using the chain rule,

$$\frac{df}{dx} = \frac{df}{dz} \times \frac{dz}{dy} \times \frac{dy}{dx}$$

$$\Rightarrow \frac{df}{dx} = \frac{d}{dz}\left(e^{-z/2}\right) \times \frac{d}{dy}\left(y^T S^{-1} y\right) \times \frac{d}{dx}(x - \mu)$$

so,

$$\rightarrow \frac{d}{dz}\left(e^{-z/2}\right) = -\frac{1}{2} e^{-z/2}.$$

$$\rightarrow \frac{d}{dy}\left(y^T S^{-1} y\right) = \lim_{h \to 0} \frac{g(y+h) - g(y)}{h}.$$

$$= \lim_{h \to 0} \frac{(y^T + h) S^{-1} (y+h) - y^T S^{-1} y}{h}$$

$$= \lim_{h \to 0} \frac{\cancel{y^T S^{-1} y} + y^T S^{-1} h + h S^{-1} y + h S^{-1} h - \cancel{y^T S^{-1} y}}{h}$$

$$= \lim_{x \to 0} \frac{h(y^T S^{-1} + S^{-1} y + S^{-1} h)}{h}$$

$$= \lim_{\lambda \to 0} \left(y^T s^{-1} + s^{-1} y + s^{-1} h \right).$$

$$= \lim_{h \to 0} y^T s^{-1} + s^{-1} y + s^{-1} h$$

$$= y^T s^{-1} + s^{-1} y$$

$$= \frac{y^T}{s} + \frac{y}{s}$$

$\longrightarrow$. $\frac{d}{dx}(x - \mu) = 1$

so, $$\frac{df}{dx} = \frac{df}{dz} \times \frac{dz}{dy} \times \frac{dy}{dx}$$

$$= \left(-\tfrac{1}{2} e^{-z/2}\right)\left(y^T s^{-1} + s^{-1} y\right) \times 1$$

so, the derivative is $-\frac{1}{2} e^{-z/2} \left(y^T s^{-1} + s^{-1} y\right)$.