PART III

EXERCISE 5.1

$$f(x) = \log(x^4)\sin(x^3)$$

$$f'(x) = \frac{4x^3}{x^4} \cdot \sin(x^3) + \log(x^4) \cdot \cos(x^3) \cdot 3x^2$$

$$= \frac{4}{x}\sin(x^3) + 3x^2\log(x^4)\cos(x^3)$$

EXERCISE 5.2

$$f(x) = \frac{1}{1+\exp(-x)}$$

$$f'(x) = -\frac{\exp(-x)}{(1+\exp(-x))^2} \cdot (-1) = \frac{\exp(-x)}{(1+\exp(-x))^2}$$

EXERCISE 5.3

$$f(x) = \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$f'(x) = \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \cdot \left(-\frac{1}{2\sigma^2}\right) \cdot 2(x-\mu)$$

$$= \frac{\mu - x}{\sigma^2} \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot (x-\mu)^2\right)$$

EXERCISE 5.5

$$J = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_m} \end{bmatrix}$$

- $f_1(x) = \sin(x_1)\cos(x_2) \qquad x \in \mathbb{R}^2$

$$\frac{\partial f_1}{\partial x_1} = \cos(x_1)\cos(x_2) \qquad \frac{\partial f_1}{\partial x_2} = -\sin(x_1)\sin(x_2)$$

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \cos(x_1)\cos(x_2) & -\sin(x_1)\sin(x_2) \end{bmatrix} \in \mathbb{R}^{1\times 2}$$

- $f_2(x,y) = x^T y \qquad x, y \in \mathbb{R}^m$

$$\frac{\partial f_2}{\partial x} = y^T \in \mathbb{R}^m \qquad \frac{\partial f_2}{\partial y} = x^T \in \mathbb{R}^m$$

$$J = \begin{bmatrix} \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} y^T & x^T \end{bmatrix} \in \mathbb{R}^{1\times 2m}$$

• $f_3(x) = \underbrace{x x^T}_{\in \mathbb{R}^{m \times m}}$ $\quad x \in \mathbb{R}^m$

Each element is given by $f_{ij} = x_i \cdot x_j$, so we will have:

$$\frac{\partial f_{ij}}{\partial x_k} = \begin{cases} x_j & \text{if } k = i \\ x_i & \text{if } k = j \\ 0 & \text{otherwise} \end{cases}$$

Since each element of f_{ij} depends on m variables, $\frac{\partial f}{\partial x}$ will be $m \times m$

$$f(x) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} [x_1 \; x_2 \ldots x_m] = \begin{bmatrix} x_1 x_1 & x_1 x_2 & \cdots & x_1 x_m \\ x_2 x_1 & x_2 x_2 & \cdots & x_2 x_m \\ \vdots & \vdots & & \vdots \\ x_m x_1 & x_m x_2 & \cdots & x_m x_m \end{bmatrix}$$

$$J = \begin{bmatrix} \frac{df_1}{dx_1} & \cdots & \frac{df_m}{dx_1} \\ \frac{df_2}{dx_2} & \cdots & \frac{df_m}{dx_2} \\ \vdots & & \\ \frac{df}{dx} & \cdots & \frac{df_m}{dx_m} \end{bmatrix} = \begin{bmatrix} 2x_1 & x_2 & \cdots & x_m \\ x_1 & 2x_2 & \cdots & x_m \\ \vdots & \vdots & & \vdots \\ x_1 & x_2 & \cdots & 2x_m \end{bmatrix}$$

EXERCISE 5.6

$f(t) = \sin(\log(t^T t)) \qquad t \in \mathbb{R}^D$

$$\frac{df}{dt} = 2t^T \cdot \frac{1}{t^T t} \cdot \cos(\log(t^T t))$$

$g(X) = \mathrm{tr}(AXB) \qquad A \in \mathbb{R}^{D \times E}, X \in \mathbb{R}^{E \times F}, B \in \mathbb{R}^{F \times D}$

$\mathrm{tr}(M) = \sum_{i=1}^{D} M_{ii}$ $\quad$ sum over the diagonal elements of the matrix $M \in \mathbb{R}^{D \times D}$

computing the matrix product:

$$g(X) = \sum_{k=1}^{D} \sum_{j=1}^{F} \sum_{i=1}^{E} a_{ki} x_{ij} b_{jk}$$

$$\frac{\partial g}{\partial x_{ij}} = \sum_{k=1}^{D} b_{jk} a_{ki} \leftarrow (i,j)\text{-th entry of } \frac{dg}{dX}$$

$$\rightarrow \frac{dg}{dX} = A^T B^T$$

EXERCISE 5.7 a. $f(z) = \log(1+z) \qquad z = x^T x \qquad x \in \mathbb{R}^D$

For the chain rule: $\dfrac{df}{dx} = \dfrac{df}{dz} \dfrac{dz}{dx}$

$$\left.\begin{aligned} \frac{dz}{dx} &= 2x^T \\ \frac{df}{dz} &= \frac{1}{1+z} \end{aligned}\right\} \frac{df}{dx} = 2x^T \cdot \frac{1}{1+z} = \frac{2x^T}{1+x^T x}$$

b. $f(z) = \sin(z) \qquad z = Ax + b \qquad A \in \mathbb{R}^{E \times D}, x \in \mathbb{R}^D, b \in \mathbb{R}^E$

The chain rule: $\dfrac{df}{dx} = \dfrac{df}{dz} \cdot \dfrac{dz}{dx}$

$$\left.\begin{aligned} \frac{df}{dz} &= \cos(z) \\ \frac{dz}{dx} &= A \end{aligned}\right\} \frac{df}{dx} = \cos(z) \cdot A$$

EXERCISE 5.8

a. $f(z) = e^{-\frac{1}{2}z} \qquad \underbrace{z = g(y) = y^T S^{-1} y}_{\text{Quadratic form}} \qquad y = h(x) = x - \mu$

$x, \mu \in \mathbb{R}^D, \quad S \in \mathbb{R}^{D \times D}$

Applying the chain rule: $\dfrac{df}{dz} = -\dfrac{1}{2} e^{-\frac{1}{2}z}$ (scalar)

$\dfrac{dz}{dy} = (S^{-1} + S^T) y$ (vector $\in \mathbb{R}^D$, same dim as y)

$\dfrac{dy}{dx} = 1$

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

$$= -\frac{1}{2} e^{-\frac{1}{2}z} \cdot (S^{-1} + S^T) y \cdot 1$$

$\dfrac{df}{dx}$ will have the same dimension as $x \rightarrow \in \mathbb{R}^D$

b. $f(x) = \mathrm{tr}(xx^T + \sigma^2 I) \qquad x \in \mathbb{R}^D$

Break down the outer product:

$$xx^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} [x_1 \; x_2 \ldots x_D] = \begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_D \\ \vdots & \vdots & & \vdots \\ x_Dx_1 & x_Dx_2 & \cdots & x_Dx_D \end{bmatrix}$$

$$\frac{df}{dx} = \frac{d}{dx}\mathrm{tr}(xx^T + \sigma^2 I) \qquad x \in \mathbb{R}^D$$

Using the cyclic property of the trace, we get:

$$\frac{df}{dx} = \underbrace{\frac{d}{dx}\mathrm{tr}(xx^T)}_{A} + \underbrace{\frac{d}{dx}\mathrm{tr}(\sigma^2 I)}_{B}$$

$$A = \frac{d}{dx}\mathrm{tr}(xx^T) = \frac{d}{dx}(x_1x_1 + x_2x_2 + x_3x_3 + \ldots + x_Dx_D)$$

$$= \begin{bmatrix} \frac{d(x_1x_1)}{dx_1} \\ \frac{d(x_2x_2)}{dx_2} \\ \vdots \\ \frac{d(x_Dx_D)}{dx_D} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \\ \vdots \\ 2x_D \end{bmatrix} = 2x$$

$$B = \frac{d}{dx}\mathrm{tr}(\sigma^2 I) = 0$$

$$\longrightarrow \frac{df}{dx} = \frac{d}{dx}\mathrm{tr}(xx^T) + \frac{d}{dx}\mathrm{tr}(\sigma^2 I) = 2x$$

c. $f = \tanh(z) \in \mathbb{R}^M \qquad z = Ax + b,\; x \in \mathbb{R}^N,\; A \in \mathbb{R}^{M\times N},\; b \in \mathbb{R}^M$

Applying chain rule:

$$\left.\begin{aligned} \frac{df}{dz} &= \mathrm{sech}^2(z) \in \mathbb{R}^{M\times N} \\ \frac{dz}{dx} &= A \qquad \in \mathbb{R}^{M\times N} \end{aligned}\right\} \quad \frac{df}{dx} = \mathrm{sech}^2(z)\cdot A \qquad \in \mathbb{R}^{M\times N}$$

EXERCISE 5.9

$$\begin{cases} g(x,z,\upsilon) = \log p(x,z) - \log q(z,\upsilon) \\ z = t(\varepsilon,\upsilon) \end{cases}$$

$$x \in \mathbb{R}^D, \quad z \in \mathbb{R}^E, \quad \upsilon \in \mathbb{R}^F, \quad \varepsilon \in \mathbb{R}^G$$

$$g(x,z,\upsilon) = \log p(x, t(\varepsilon,\upsilon)) - \log q(t(\varepsilon,\upsilon),\upsilon)$$

$$\frac{d}{d\upsilon} g(x,\varepsilon,\upsilon) = \frac{d}{d\upsilon}(\log p(x,t(\varepsilon,\upsilon)) - \frac{d}{d\upsilon}(\log q(t(\varepsilon,\upsilon),\upsilon)$$

$$= \frac{dt(\varepsilon,\upsilon)}{d\upsilon} \frac{p'(x,t(\varepsilon,\upsilon))}{p(x,t(\varepsilon,\upsilon))} - \frac{dt(\varepsilon,\upsilon)}{d\upsilon} \cdot \frac{q'(t(\varepsilon,\upsilon),\upsilon)}{q(t(\varepsilon,\upsilon),\upsilon)}$$

$$= \frac{t'(\varepsilon,\upsilon)p'(x,t(\varepsilon,\upsilon))}{p(x,t(\varepsilon,\upsilon))} - \frac{t'(\varepsilon,\upsilon)q'(t(\varepsilon,\upsilon),\upsilon)}{q(t(\varepsilon,\upsilon),\upsilon)}$$