5.8 Compute the derivatives df/dx of the following functions. Describe your steps in detail.

a. Use the chain rule. Provide the dimensions of every single partial derivative.

$$f(z) = exp(-\frac{1}{2}z)$$

$$z = g(\mathbf{y}) = \mathbf{y}^T\mathbf{S}^{-1}\mathbf{y}$$

$$\mathbf{y} = h(\mathbf{x}) = \mathbf{x} - \boldsymbol{\mu}$$

where $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^D, \mathbf{S} \in \mathbb{R}^{D\times D}$.

Solution.

$$\frac{df}{dx} = \frac{\partial f}{\partial z} * \frac{\partial z}{\partial y} * \frac{\partial y}{\partial x}$$

$$\frac{\partial y}{\partial x} = \frac{\partial h}{\partial x} = \mathbf{I}_D. \text{ Dimensions are DxD.}$$

$$\frac{\partial z}{\partial y} = \frac{\partial g}{\partial y} = \mathbf{y}^T(\mathbf{S}^{-1} + (\mathbf{S}^{-1})^T). \text{ Using Identity 5.107.}$$

$$= (\mathbf{x} - \boldsymbol{\mu})^T(\mathbf{S}^{-1} + (\mathbf{S}^{-1})^T). \text{ Dimensions are 1xD.}$$

$$\frac{\partial f}{\partial z} = -\frac{1}{2}exp\left(-\frac{1}{2}z\right) = -\frac{1}{2}exp\left(-\frac{1}{2}(\mathbf{y}^T\mathbf{S}^{-1}\mathbf{y})\right)$$

$$= -\frac{1}{2}exp\left(-\frac{1}{2}((\mathbf{x} - \boldsymbol{\mu})^T\mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu}))\right). \text{ Dimensions are 1x1.}$$

$$\implies \frac{df}{dx} = -\frac{1}{2} exp\left(-\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\right)\left((\mathbf{x}-\boldsymbol{\mu})^T(\mathbf{S}^{-1}+(\mathbf{S}^{-1})^T)\right)\mathbf{I}_D$$

$$\implies \frac{df}{dx} = -\frac{1}{2} exp\left(-\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\right)\left((\mathbf{x}-\boldsymbol{\mu})^T(\mathbf{S}^{-1}+(\mathbf{S}^{-1})^T)\right)$$

b.

$$f(\mathbf{x}) = tr(\mathbf{x}\mathbf{x}^T + \sigma^2\mathbf{I}), \qquad \mathbf{x} \in \mathbb{R}^D$$

Here tr($\mathbf{A}$) is the trace of $\mathbf{A}$, i.e., the sum of the diagonal elements A_{ii}.

Hint: Explicitly write out the outer product.

Solution.

$$\frac{df}{d\mathbf{x}} = \frac{d(tr(\mathbf{x}\mathbf{x}^T + \sigma^2\mathbf{I}))}{d\mathbf{x}} = \frac{d(tr(\mathbf{x}\mathbf{x}^T) + tr(\sigma^2\mathbf{I}))}{d\mathbf{x}} = \frac{d(tr(\mathbf{x}\mathbf{x}^T))}{d\mathbf{x}} + \frac{d(tr(\sigma^2\mathbf{I}))}{d\mathbf{x}} = \frac{d(tr(\mathbf{x}\mathbf{x}^T))}{d\mathbf{x}}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_D \\ x_2^2 & x_2^2 & \dots & x_2x_D \\ \vdots & \vdots & \dots & \vdots \\ x_D^2 & x_Dx_2 & \dots & x_D^2 \end{bmatrix}$$

$$\implies tr(\mathbf{x}\mathbf{x}^T) = \Sigma_{i=1}^D x_i^2$$

$$\implies \frac{df}{d\mathbf{x}} = \frac{d(\Sigma_{i=1}^{D} x_i^2)}{d\mathbf{x}} = \begin{bmatrix} \frac{d(\Sigma_{i=1}^{D} x_i^2)}{d\mathbf{x}_1} & \frac{d(\Sigma_{i=1}^{D} x_i^2)}{d\mathbf{x}_2} & \dots & \frac{d(\Sigma_{i=1}^{D} x_i^2)}{d\mathbf{x}_D} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 & 2x_2 & \dots & 2x_D \end{bmatrix} = 2 \begin{bmatrix} x_1 & x_2 & \dots & x_D \end{bmatrix} = 2\mathbf{x}^T$$

c. Use the chain rule. Provide the dimensions of every single partial derivative.

You do not need to compute the product of the partial derivatives explicitly.

$$\mathbf{f} = tanh(\mathbf{z}) \in \mathbb{R}^M$$

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{b}, \qquad \mathbf{x} \in \mathbb{R}^N, \mathbf{A} \in \mathbb{R}^{M \times N}, \mathbf{b} \in \mathbb{R}^M.$$

Solution.

$$\frac{d\mathbf{f}}{d\mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{z}} * \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial(\mathbf{A}\mathbf{x} + \mathbf{b})}{\partial \mathbf{x}} = \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} + \frac{\partial \mathbf{b}}{\partial \mathbf{x}} = \mathbf{A} + \mathbf{0} = \mathbf{A}$$ As calculated in 5.7. Dimensions are MxN

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}} = \frac{\partial \tanh(\mathbf{z})}{\partial \mathbf{z}}$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_M \end{bmatrix}$$

$$\implies tanh(\mathbf{z}) = \begin{bmatrix} tanh(z_1) \\ tanh(z_2) \\ \vdots \\ tanh(z_M) \end{bmatrix}$$

$$\implies \frac{\partial \mathbf{f}}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial tanh(z_1)}{\partial z_1} & \frac{\partial tanh(z_1)}{\partial z_2} & \dots & \frac{\partial tanh(z_1)}{\partial z_M} \\ \frac{\partial tanh(z_2)}{\partial z_1} & \frac{\partial tanh(z_2)}{\partial z_2} & \dots & \frac{\partial tanh(z_2)}{\partial z_M} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial tanh(z_M)}{\partial z_1} & \frac{\partial tanh(z_M)}{\partial z_2} & \dots & \frac{\partial tanh(z_M)}{\partial z_M} \end{bmatrix}$$

$$= \begin{bmatrix} 1 - tanh^2(x_1) & 0 & \dots & 0 \\ 0 & 1 - tanh^2(x_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 - tanh^2(x_M) \end{bmatrix}$$

$\frac{\partial \mathbf{f}}{\partial \mathbf{z}}$ has dimensions MxM.