5. Vector Calculus (Exercises Only)

Exercises

5.1

- Compute the derivative $f'(x)$ for:

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

- Product rule:

$$f'(x) = g'(x)h(x) + g(x)h'(x)$$

$$g'(x) = \frac{4}{x}$$

$$h'(x) = 3x^2\cos(x^3)$$

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

$$f'(x) = \frac{4}{x}\sin(x^3) + 12x^2\log(x)\cos(x^3)$$

5.2

- Compute the derivative $f'(x)$ of the logistic sigmoid:

First, let's rewrite this fraction as a negative exponent:

$$\sigma = (1 + e^{-z})^{-1}$$

No we can apply the chain rule:

This is the "inner function":

$$g(z) = 1 + e^{-z}$$

This is the "outer function":

$$f(z) = z^{-1}$$

So first we find the derivative of the "inner function":

$$g'(z) = -e^{-z}$$

Now we find the derivative of the "outer function" if the "inner function" were just a variable:

$$f'(z) = -(1 + e^{-z})^{-2}$$

Now, as per the chain rule, we multiply them together:

$$\sigma'(z) = g'z * f'(z)$$

$$\sigma'(z) = -e^{-z} - (1 + e^{-z})^{-2}$$

Now, let's convert our negative exponent back to a fraction:

$$\sigma'(z) = \frac{e^{-z}}{(1 + e^{-z})^2}$$

Let's factor this into two pieces:

$$\sigma'(z) = \frac{1}{1 + e^{-z}} \cdot \frac{e^{-z}}{1 + e^{-z}}$$

We can rewrite the numerator of the second part to look like this:

$$\sigma'(z) = \frac{1}{1 + e^{-z}} \cdot \frac{(1 + e^{-z}) - 1}{1 + e^{-z}}$$

All we did was just add 1 to the numerator and then subtract 1. This does not change the value at all, but it allows us to factor like this:

$$\sigma'(z) = \frac{1}{1 + e^{-z}} \cdot \frac{1 + e^{-z}}{1 + e^{-z}} - \frac{1}{1 + e^{-z}}$$

Simplifying further gives us:

$$\sigma'(z) = \frac{1}{1 + e^{-z}} \cdot 1 - \frac{1}{1 + e^{-z}}$$

We know that: $\sigma(z) = \dfrac{1}{1 + e^{-z}}$, so we can substitute into our equation which finally gives us:

$$\sigma'(z) = \sigma'(z) \cdot (1 - \sigma'(z))$$

5.3

- Compute the derivative $f'(x)$ of the function

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

- where $\mu, \sigma \in \mathbb{R}$ are constants
- Chain rule:

$$f'(x) = g'(x) \cdot h'(g(x))$$

$$g'(x) = -\frac{2}{2\sigma^2}(x-\mu) = -\frac{(x-\mu)}{\sigma^2}$$

$$h'(g(x)) = \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$$

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

5.4

- Compute the Taylor polynomials $T_n, n = 0, \ldots, 5$ of

$$f(x) = \sin(x) + \cos(x)$$

at

$$x_0 = 0$$

- I will compute each term in the Taylor polynomial $\hat{T}_i$ individually, then combine them together at the end such that:

$$T_5 = \sum_{i=0}^{5} \hat{T}_i$$

$$\hat{T}_0 = \frac{D_x^0 f(0)}{0!}\delta^0 = f(0) = 1$$

$$\hat{T}_1 = \frac{D_x^1 f(0)}{1!}\delta^1 = \cos(0) - \sin(0)x = x$$

$$\hat{T}_2 = \frac{D_x^2 f(0)}{2!}\delta^2 = \frac{1}{2} - \sin(0) - \cos(0)x^2 = -\frac{1}{2}x^2$$

$$\hat{T}_3 = \frac{D_x^3 f(0)}{3!} = \frac{1}{6} - \cos(0) + \sin(0)x^3 = -\frac{1}{6}x^3$$

$$\hat{T}_4 = \frac{D_x^4 f(0)}{4!} = \frac{1}{24}\sin(0) + \cos(0)x^4 = \frac{1}{24}x^4$$

$$\hat{T}_5 = \frac{D_x^5 f(0)}{5!} = \frac{1}{120}\cos(0) - \sin(0)x^5 = \frac{1}{120}x^5$$

$$T_5 = 1 + x - \frac{1}{2}x^2 - \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5$$

5.5

- Consider the following functions:

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

$$f_2(x) = xTy, \quad x, y \in \mathbb{R}^n$$

$$f_3(x) = xx^T, \quad x \in \mathbb{R}^n$$

a. What are the dimensions of $\dfrac{\partial f_i}{\partial x}$?

$$f_1 : \frac{\partial f_1}{\partial x} \in \mathbb{R}^{1\times 2}$$

$$f_2 : \frac{\partial f_2}{\partial x} \in \mathbb{R}^{1\times n}$$

$$f_3 : \frac{\partial f_3}{\partial x} \in \mathbb{R}^{n\times n\times n}$$

b. Compute the Jacobians:

$$f_1 : \frac{\partial f_1}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \end{bmatrix} = [\cos(x_1)\cos(x_2) \quad -\sin(x_1)\sin(x_2)]$$

$$f_2 : \frac{\partial f_2}{\partial x} = \begin{bmatrix} \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \end{bmatrix} = y^T$$

$$f_3 : \frac{\partial f_3}{\partial x} = \begin{bmatrix} \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \dots & \frac{\partial f_3}{\partial x_n} \end{bmatrix}$$

- In this case each of $\dfrac{\partial f_3}{\partial x_i}$ are matrices, so this will be an $n \times n \times n$ tensor, as we said in part a.

$$\frac{\partial f_3}{\partial x} = \begin{bmatrix} \begin{bmatrix} 2x_1 & x_2 & \dots & x_n \\ x_2 & 0 & \dots & 0 \\ \vdots & 0 & \dots & 0 \\ x_n & 0 & \dots & 0 \end{bmatrix} & \begin{bmatrix} 0 & x_2 & \dots & 0 \\ x_1 & 2x_2 & \dots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ 0 & x_n & 0 & 0 \end{bmatrix} & \dots & \begin{bmatrix} 0 & 0 & \dots & x_1 \\ 0 & 0 & \dots & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1 & x_2 & \dots & 2x_n \end{bmatrix} \end{bmatrix}$$

5.6

- Differentiate f with respect to t and g with respect to X where:

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

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

a.

$$f(t) = \sin(\log(t^T t))$$

- I will break this into smaller functions, then use the chain rule to easily compute the derivative

$$a = t^T t$$

$$b = \log(a)$$

$$f = \sin(b)$$

$$\frac{\partial a}{\partial t} = 2t^T$$

$$\frac{\partial b}{\partial a} = \frac{1}{a}$$

$$\frac{\partial f}{\partial b} = \cos(b)$$

- Now we can use the chain rule:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial t}$$

$$\frac{\partial f}{\partial t} = \cos(\log(t^T t))\frac{2t^T}{t^T t}$$

b.

$$g(X) = \mathrm{tr}(AXB)$$

Using the property of the trace that allows us to "rotate" the matrix multiplications:

$$g(X) = \mathrm{tr}(AXB) = \mathrm{tr}(BAX)$$

Now, we also know that:

$$\frac{\partial}{\partial X}\mathrm{tr}(CX) = X^T$$

- Therefore:

$$\frac{\partial g}{\partial X} = (BA)^T = A^T B^T$$

5.7

- Compute the derivatives $\frac{df}{dx}$ of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a.

$$f(z) = \log(1+z), \quad z = x^T x, \quad x \in \mathbb{R}^D$$

$$\frac{\partial z}{\partial x} \in \mathbb{R}^{1\times D}$$

$$\frac{\partial z}{\partial x} = 2x^T$$

$$\frac{\partial f}{\partial z} \in \mathbb{R}$$

$$\frac{\partial f}{\partial z} = \frac{1}{1+z}$$

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{1\times D}$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = \frac{1}{1+z}2x^T = \frac{1}{1+x^Tx}2x^T$$

b.

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

$$\frac{\partial z}{\partial x} \in \mathbb{R}^{ExD}$$

$$\frac{\partial z}{\partial x} = A$$

$$\frac{\partial f}{\partial z} \in \mathbb{R}^{E\times E}$$

$$\frac{\partial f}{\partial z} = \begin{bmatrix} \cos(z_1) & 0 & \dots & 0 \\ 0 & \cos(z_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cos(z_E) \end{bmatrix}$$

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{E\times D}$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = \begin{bmatrix} \cos(z_1) & 0 & \dots & 0 \\ 0 & \cos(z_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cos(z_E) \end{bmatrix} A$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \cos(z_1)a_{11} & \cos(z_1)a_{12} & \dots & \cos(z_1)a_{1D} \\ \cos(z_2)a_{21} & \cos(z_2)a_{22} & \dots & \cos(z_2)a_{2D} \\ \vdots & \vdots & \vdots & \vdots \\ \cos(z_E)a_{E1} & \cos(z_E)a_{E2} & \dots & \cos(z_E)a_{ED} \end{bmatrix}$$

5.8

- Compute the derivatives of $\frac{df}{dx}$ of the following functions. Describe your steps in detail.

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

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

$$z = g(y) = y^TS^{-1}y$$

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

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

$$\frac{\partial y}{\partial x} \in \mathbb{R}^{D \times D}$$

$$\frac{\partial y}{\partial x} = I$$

$$\frac{\partial z}{\partial y} \in \mathbb{R}^{1 \times D}$$

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

$$\frac{\partial f}{\partial z} \in \mathbb{R}$$

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

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{1 \times D}$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial y}\frac{\partial y}{\partial x} = \exp(-\frac{1}{2}z)y^T(S^{-1} + (S^{-1})^T)I = \exp(-\frac{1}{2}y^T S^{-1} y)y^T(S^{-1} + (S^{-1})^T)$$

b.

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

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

$$xx^T + \sigma^2 I = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_D \\ x_1x_2 & x_2^2 & \dots & x_2x_D \\ \vdots & \vdots & \vdots & \vdots \\ x_1x_D & x_2x_D & \dots & x_D^2 \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

$$xx^T + \sigma^2 I == \begin{bmatrix} x_1^2 + \sigma^2 & x_1x_2 & \dots & x_1x_D \\ x_1x_2 & x_2^2 + \sigma^2 & \dots & x_2x_D \\ \vdots & \vdots & \vdots & \vdots \\ x_1x_D & x_2x_D & \dots & x_D^2 + \sigma^2 \end{bmatrix}$$

$$\mathrm{t}r(xx^T + \sigma^2 I) = \sum_{i=0}^{D} x_i^2 + \sigma^2$$

$$\frac{\partial}{\partial x}\left(\sum_{i=0}^{D} x_i^2 + \sigma^2\right) \in \mathbb{R}^{1\times D}$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_D} \end{bmatrix} = [2x_1 \quad 2x_2 \quad \dots \quad 2x_D] = 2x^T$$

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.

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

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

- This is basically a linear layer and a tanh nonlinearity from a standard feedforward neural net!

$$\frac{\partial z}{\partial x} \in \mathbb{R}^{M\times N}$$

$$\frac{\partial z}{\partial x} = A$$

$$\frac{\partial f}{\partial z} \in \mathbb{R}^{M\times M}$$

$$\frac{\partial f}{\partial z} = \mathrm{diag}(1-\tanh^2(z_1), 1-\tanh^2(z_2), \dots 1-\tanh^2(z_M))$$

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{M\times N}$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = \mathrm{diag}(1-\tanh^2(z_1), 1-\tanh^2(z_2), \dots 1-\tanh^2(z_M))A$$

5.9

- We define:

$$g(z,\nu) := \log p(x,z) - \log q(z,\nu)$$

$$z := t(\epsilon, \nu)$$

- for differentiable functions p, q, t. By using the chain rule, compute the gradient:

$$\frac{d}{d\nu} g(z,\nu)$$

$$\frac{\partial g}{\partial \nu} = \frac{1}{p(x, t(\epsilon, \nu))} \frac{\partial p}{\partial z} \frac{\partial z}{\partial \nu} - \frac{1}{q(t(\epsilon, \nu), \nu)} \frac{\partial q}{\partial \nu} + \frac{\partial q}{\partial z} \frac{\partial z}{\partial v}$$