Exercise 2.1 - Linear Algebra (0.5+1+2 points)

a) Compute the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$

b) Given a matrix A for which an inverse exists, find the relationship between the eigenvalues of A and A^{-1}.

c) Show that if λ is an eigenvalue of AB, then it is also an eigenvalue of BA where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times n}$.

a) We take the characteristic polynomial of A

$$\chi_A(\lambda) = \det(A - \lambda I) = (4-\lambda)(3-\lambda) - 2\cdot 1$$
$$= \lambda^2 - 7\lambda + 10$$

Solving the quadratic equation yields $\lambda_1 = 2$ and $\lambda_2 = 5$.
The corresponding eigenvectors are given by

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

b) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then $A^{-1} = A$. Let v be an eigenvector of an arbitrary invertible matrix B. Then:

$$Bv = \lambda v \iff B^{-1}Bv = \lambda Bv$$
$$\iff v = \lambda Bv$$
$$\iff \frac{1}{\lambda} = Bv$$

hence, if λ is an eigenvalue of B, then $\frac{1}{\lambda}$ is an eigenvalue of B^{-1} if $\lambda \neq 0$. If $\lambda = 0$, then λ is also an eigenvalue of B^{-1} with eigenvector Bv. $\blacksquare$

c) Let λ be an eigenvalue of AB with eigenvector v.
We have

$$ABv = \lambda v \iff BA(Bv) = \lambda(Bv) \quad | \; v' = Bv \in \mathbb{R}^n$$
$$\iff BAv' = \lambda v'$$

$\blacksquare$

Exercise 2.2 - Matrix Calculus (0.5+1+1+1 points)

In this lecture we will often compute the derivatives of multivariate functions and matrix valued functions. Let $f : \mathbb{R}^n \to \mathbb{R}$; $w, x, c \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$. Prove that the following rules hold.

a) $f(x) = w^T x$, then $\nabla_x f(x) = w$

b) $f(x) = x^T A x$, then $\nabla_x f(x) = Ax + A^T x$

c) $f(x) = \|Bx\|_2^2$, then $\nabla_x f(x) = 2B^T B x$

d) $f(x) = \|Bx - c\|_2^2$, then $\nabla_x f(x) = 2B^T(Bx - c)$

a) $\nabla_X f(X) = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]^T$

$= \left[w_1, \dots, w_n \right]^T$

$= w$

b) We have $x^T A x = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} x_j$, whose partial derivative is

$\frac{\partial x^T A x}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{i,j=1}^{n} x_i a_{ij} x_j = \sum_{j \neq 1}^{n} a_{1j} x_j + a_{j1} x_j + 2a_{11} x_1 = (Ax)_1 + (A^T x)_1$

$\nabla_x f(x) = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]^T$

$= \left[\sum_{j \neq 1}^{n} a_{1j} x_j + 2 a_{11} x_1, \dots, \sum_{j \neq n}^{n} a_{nj} x_j + 2 a_{nn} x_n \right]^T$

$= (Ax + A^T x)$

c) We can rewrite the function as

$$f(x) = \| Y \|_2^2 \text{ with } Y = Bx \in \mathbb{R}^n.$$

We can then apply the chain-rule to get the derivative of f w.r.t x. We have $\|Y\|_2^2 = Y^T Y$

$$\frac{\partial f}{\partial x} = \underbrace{\frac{\partial Y}{\partial x}}_{\in \mathbb{R}^{n \times n}} \cdot \underbrace{\frac{\partial Y^T Y}{\partial Y}}_{\in \mathbb{R}^{n \times 1}} \in \mathbb{R}^{n \times 1}$$

It can be shown in an analogous way to a) that

$$\frac{\partial Y^T Y}{\partial Y} = 2Y$$

For the second term of the chain rule, we have

$$Bx = \begin{pmatrix} \sum_{i=1}^{n} b_{1i} x_i \\ \vdots \\ \sum_{i=1}^{n} b_{ni} x_i \end{pmatrix} \in \mathbb{R}^n$$

Which means that $\frac{\partial Y}{\partial x}$ is an $n \times n$ matrix, with

$$\left(\frac{\partial Y}{\partial x}\right)_{ij} = \frac{\partial Y_i}{\partial x_j} = b_{ji}$$

Therefore $\frac{\partial Bx}{\partial x} = B^T$ and we have

$$\nabla_x f(x) = B^T \cdot 2Y = B^T 2Bx = 2B^T Bx$$

d) We can use the chain-rule again, but with $Y = Bx - c$. We still have $\frac{\partial Y}{\partial x} = B^T$, because c is a constant. The gradient is then given by

$$\nabla f_x(x) = 2B^T(Bx - c)$$