Assignment 1

Due Tuesday, October 5, 2021 at 11:59pm

This assignment is to be done individually.

Important Note: The university policy on academic dishonesty (cheating) will be taken very seriously in this course. You may not provide or use any solution, in whole or in part, to or by another student.

You are encouraged to discuss the concepts involved in the questions with other students. If you are in doubt as to what constitutes acceptable discussion, please ask! Further, please take advantage of office hours offered by the instructor and the TA if you are having difficulties with this assignment.

DO NOT:

- Give/receive code or proofs to/from other students
- Use Google to find solutions for assignment

DO:

- Meet with other students to discuss assignment (it is best not to take any notes during such meetings, and to re-work assignment on your own)
- Use online resources (e.g. Wikipedia) to understand the concepts needed to solve the assignment.

Submitting Your Assignment

The assignment must be submitted online on Canvas. You must submit a report in **PDF format**. You may typeset your assignment in LaTeX or Word or submit neatly handwritten and scanned solutions. We will not be able to give credit to solutions that are not legible.

1 Convexity

For any $\overrightarrow{x}, \overrightarrow{y} \in \mathbb{R}^n$ and any $t \in [0,1]$, a function f is said to be convex if it satisfies any of these conditions:

- $f(t\overrightarrow{x} + (1-t)\overrightarrow{y}) \leq tf(\overrightarrow{x}) + (1-t)f(\overrightarrow{y})$
- If f is differentiable: $f(\overrightarrow{y}) \geq f(\overrightarrow{x}) + \nabla f(\overrightarrow{x}) \cdot (\overrightarrow{y} - \overrightarrow{x})$
- If f is twice differentiable: $\mathrm{H}f(\overrightarrow{x}) \geq 0$

a) Given $x \in \mathbb{R}$ and only using the definition of convex functions given above, prove that the rectified linear unit function, $\mathrm{ReLU}(x) := \max(x, 0)$, is convex.

b) Given a $A \in \mathbb{R}^{n \times n}$, $\overrightarrow{x} \in \mathbb{R}^n$, $\overrightarrow{b} \in \mathbb{R}^n$, and $\lambda \geq 0$, prove that $f(\overrightarrow{x}) = \left\| A\overrightarrow{x} + \overrightarrow{b} \right\|_2 + \lambda \left\| \overrightarrow{x} \right\|_1$ is convex. For this part, you may use the following properties of convex functions:

   - $\sum_i w_i f_i(\overrightarrow{x})$ is convex if f_i are convex and $w_i \geq 0$
   - For any $A \in \mathbb{R}^{n \times n}$ and $\overrightarrow{b} \in \mathbb{R}^n$, $g(\overrightarrow{x}) = f(A\overrightarrow{x} + \overrightarrow{b})$ is convex if f is convex
   - $g(f(\overrightarrow{x}))$ is convex if f is convex and g is convex and non-decreasing

c) Given $x \in \mathbb{R}$, prove that the logistic function $f(x) = \frac{1}{1+e^{-x}}$ is neither convex nor concave (i.e. $-f(x)$ is also not convex).

Part A

$$ReLU(x) = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases}$$

We know by ReLU's graph that it is convex, but we can use the first condition, or third for that matter, of convexity to prove that ReLU is infact a convex function.
if we take ReLU of x + y it will be like taking individual ReLU's of x and y. And if we multiply a scaler with x it will just carry it to the outside of ReLU.

$$ReLU(x+y) = ReLU(x) + ReLU(y)$$

$$ReLU(tx) = tReLU(x)$$

So, if we multiply x with a scaler t, $t \in [0,1]$ and y with (1-t), we get:

$$ReLU(tx + (1-t)y) = tReLU(x) + (1-t)ReLU(y)$$

which satisfies the first condition of convexity ($f(t\overrightarrow{x} + (1-t)\overrightarrow{y}) \leq tf(\overrightarrow{x}) + (1-t)f(\overrightarrow{y})$).
And if we take the second derivative of ReLU(x) it gives us 0 which satisfies the third condition of convexity.

Part B

For this part I will divide the equation into two parts, which are:

$$u(\overrightarrow{x}) = \left\| A\overrightarrow{x} + \overrightarrow{b} \right\|_2$$

$$v(\overrightarrow{x}) = \lambda \left\| \overrightarrow{x} \right\|_1$$

And I will be proving that they are convex, then will apply the condition 1 ($\sum_i w_i f_i(\overrightarrow{x})$ is convex if f_i are convex and $w_i \geq 0$).
Let's start with proving convexity of $u(\overrightarrow{x})$, Say if $u(x) = \| x\|_2$ and we add another variable, say y, it will become $u(x+y) = \| x+y\|_2$. Similarly multiplying it with a constant, t, will result in $u(tx) = \| tx\|_2$ [t can be moved outside of the norm]. Now I will perform a series of operations like these.

$$u(x) = \| x\|_2$$

Using a constant $t \in [0,1]$.

$$u(tx) = \| tx\|_2$$

Adding another variable y multiplied with the constant $(1-t)$.

$$u(tx + (1-t)y) = \| tx + (1-t)y\|_2$$

Now applying **Triangle Inequality of Norms**.

$$\| tx + (1-t)y\|_2 \leq \| tx\|_2 + \|(1-t)y\|_2$$

Constants can be moved outside of the norm.

$$\| tx + (1-t)y\|_2 \leq t\| x\|_2 + (1-t)\|y\|_2$$

If we express it in terms of functions it can be seen that it is exactly the same as the 1^{st} condition of convexity.

$$u(tx + (1-t)y) \leq tu(x) + (1-t)u(y)$$

Hence the first part of the original equaltion is convex. The second part can be proved just like this. Now for the proof of convexity of L-1 Norm, $v(\overrightarrow{x})$.

$$v(x) = \| x\|_1$$

Using a constant $t \in [0, 1]$.

$$v(tx) = \| tx\|_1$$

Adding another variable y multiplied with the constant $(1 - t)$.

$$v(tx + (1 - t)y) = \| tx + (1 - t)y\|_1$$

Now applying **Triangle Inequality of Norms**.

$$\| tx + (1 - t)y\|_1 \leq \| tx\|_1 + \|(1 - t)y\|_1$$

Constants can be moved outside of the norm.

$$\| tx + (1 - t)y\|_1 \leq t \| x\|_1 + (1 - t) \|y\|_1$$

If we again express it in terms of functions it can be seen that it is exactly the same as the 1^{st} condition of convexity.

$$v(tx + (1 - t)y) \leq tv(x) + (1 - t)v(y)$$

Which proves that both of the parts, u and v, are convex. Now if we use the condition 1 ($\sum_i w_i f_i(\overrightarrow{x})$ is convex if f_i are convex and $w_i \geq 0$) to prove that the equation $f(x) = u(x) = v(x)$ is also a convex. The condition 1 says that the functions, u(x) and v(x), need to be convex and the $w_i \geq 0$ ($\lambda \geq 0$). Both of the conditions are satisfied in this case. Hence it can be said that the equation is convex.

Part C

By looking at the graph we can already see that the Sigmoid Function is a convex function till one point and concave function from that point onwards.
For a function to be not convex (or concave for that matter) I will find the second derivative of the function and try to disprove the condition. First, I will find the first order derivative of the sigmoid function:

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

$$\frac{d}{dx}f(x) = \frac{d}{dx}(\frac{1}{1+e^{-x}})$$

$$= \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= -(1+e^{-x})^{-2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}}$$

Adding and subtracting 1.

$$= \frac{1}{1+e^{-x}}\frac{e^{-x}+1-1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\frac{(1+e^{-x})-1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}})$$

$$= \frac{1}{1+e^{-x}}(1 - \frac{1}{1+e^{-x}})$$

This can be written in the form of f(x).

$$\frac{d}{dx}f(x) = f(x)(1-f(x))$$

Now to find its derivative. The above equation can be written in the form:

$$f'(x) = f(x) - f(x)^2$$

Now finding its derivative

$$f''(x) = f'(x) - 2f(x)f'(x)$$

$$f''(x) = f'(x)(1-2f(x))$$

or we can put the equation of the f'(x) on the right hand side.

$$f''(x) = f(x)(1 - f(x))(1 - 2f(x))$$

Now if we try to find a similar relation for the concave funtion (put f(x) = -f(x)).

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

$$\frac{d}{dx}f(x) = -\frac{d}{dx}(\frac{1}{1+e^{-x}})$$

$$= -\frac{d}{dx}(1+e^{-x})^{-1}$$

$$= (1+e^{-x})^{-2}(-e^{-x})$$

$$= -\frac{e^{-x}}{(1+e^{-x})^2}$$

$$= -\frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}}$$

Adding and subtracting 1.

$$= -\frac{1}{1+e^{-x}}\frac{e^{-x}+1-1}{1+e^{-x}}$$

$$= -\frac{1}{1+e^{-x}}\frac{(1+e^{-x})-1}{1+e^{-x}}$$

$$= -\frac{1}{1+e^{-x}}(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}})$$

$$= -\frac{1}{1+e^{-x}}(1 - \frac{1}{1+e^{-x}})$$

This can be written in the form of f(x).

$$\frac{d}{dx}f(x) = -f(x)(1 - f(x))$$

Now to find its derivative. The above equation can be written in the form:

$$f'(x) = -f(x) + f(x)^2$$

Now finding its derivative

$$f''(x) = -f'(x) + 2f(x)f'(x)$$

$$f''(x) = -f'(x)(1 - 2f(x))$$

or we can put the equation of the f'(x) on the right hand side.

$$f''(x) = -(-f(x)(1 - f(x)))(1 - 2f(x))$$

Which gives us the same result as before

$$f''(x) = f(x)(1 - f(x))(1 - 2f(x))$$

Now we know that sigmoid function outputs a value from 0 to 1 ($f(x) \in [0, 1]$). So, if f(x) is anything greater than 0.5 the output will contradict the third condition of convexity (H$f(\overrightarrow{x}) \geq 0$). Say we choose f(x) = 0.6.

$$= 0.6(1 - 0.6)(1 - 1.2)$$

$$= 0.6(0.4)(-0.2)$$

$$= -(0.6)(0.4)(0.2)$$

Which contradicts the third condition. Hence, it can be said that the function is neither convex nor concave.

2 Taylor Expansion and Quadratic Form

a) Given $x \in \mathbb{R}$ and $y \in \mathbb{R}$, consider the following function

$$h(x, y) = -\cos(x^2) + e^{xy} - 2y^2$$

(i) Compute the **Gradient** and **Hessian matrix** of h

(ii) Find the **second order Taylor expansion** of h at the point $(x = x_0, y = y_0)$

(iii) Simplify the equation from (b) at the point $(x_0 = 0, y_0 = 0)$

(iv) Use result from (c), determine the definiteness for the **Hessian** of h around the point $(x_0 = 0, y_0 = 0)$.

b) (i) Let $\overrightarrow{x}, \overrightarrow{b} \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n\times n}$ be a symmetric and invertible matrix, prove the following equation:

$$(\overrightarrow{x} - M^{-1}\overrightarrow{b})^\top M(\overrightarrow{x} - M^{-1}\overrightarrow{b}) = \overrightarrow{x}^\top M\overrightarrow{x} - 2\overrightarrow{b}^\top\overrightarrow{x} + \overrightarrow{b}^\top M^{-1}\overrightarrow{b}$$

(ii) Let $\overrightarrow{x}, \overrightarrow{\mu}, \overrightarrow{\theta} \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{n\times n}$ be symmetric matrices, assume $A + B$ is invertible. Let $f(\overrightarrow{x})$ be the sum of two quadratic forms:

$$f(\overrightarrow{x}) = (\overrightarrow{x} - \overrightarrow{\mu})^\top A(\overrightarrow{x} - \overrightarrow{\mu}) + (\overrightarrow{x} - \overrightarrow{\theta})^\top B(\overrightarrow{x} - \overrightarrow{\theta})$$

Show that f can be written as a single quadratic form plus some constant term. (Hint: use (i))

Part A

Subpart 1

The partial derivatives, with respect to the variables x and y, of the original equation are:

$$\frac{\partial h}{\partial x} = 2xsin(x^2) + ye^{xy}$$
$$\frac{\partial h}{\partial y} = xe^{xy} - 4y$$

Which makes the gradient matrix/vector:

$$\overrightarrow{g} = \begin{pmatrix} 2xsin(x^2) + ye^{xy} \\ xe^{xy} - 4y \end{pmatrix}$$

Now that the first order partial derivatives have been found, we move on to the second order partial derivaties - for which the matrix is called the Hessian. The second order partial derivatives are:

$$\frac{\partial^2 h}{\partial x \partial x} = 4x^2cos(x^2) + 2sin(x^2) + y^2e^{xy}$$
$$\frac{\partial^2 h}{\partial x \partial y} = xye^{xy} + e^{xy}$$
$$\frac{\partial^2 h}{\partial y \partial x} = xye^{xy} + e^{xy}$$
$$\frac{\partial^2 h}{\partial y \partial y} = x^2e^{xy} - 4$$

The Hessian is:

$$H = \begin{pmatrix} 4x^2cos(x^2) + 2sin(x^2) + y^2e^{xy} & xye^{xy} + e^{xy} \\ xye^{xy} + e^{xy} & x^2e^{xy} - 4 \end{pmatrix}$$

Subpart 2

The second order Taylor expansion for the expression, at the point (x=x_0, y=y_0), is given by:

$$h(\overrightarrow{x_0}) + (\overrightarrow{x} - \overrightarrow{x_0})^T\overrightarrow{g} + \frac{1}{2}(\overrightarrow{x} - \overrightarrow{x_0})^T H(\overrightarrow{x} - \overrightarrow{x_0})$$

The first part gives:

$$h(\overrightarrow{x_0}) = -cos(x_0^2) + e^{x_0y_0} - 2y_0^2$$

The second part, containing the gradient vector, gives:

$$(\overrightarrow{x} - \overrightarrow{x_0})^T\overrightarrow{g} = \begin{pmatrix} x - x_0 & y - y_0 \end{pmatrix} \begin{pmatrix} 2x_0sin(x_0^2) + y_0e^{x_0y_0} \\ x_0e^{x_0y_0} - 4y_0 \end{pmatrix}$$

The third part, containing the Hessian matrix, gives:

$$\frac{1}{2}(\overrightarrow{x} - \overrightarrow{x_0})^T H(\overrightarrow{x} - \overrightarrow{x_0}) = \frac{1}{2} \begin{pmatrix} x - x_0 & y - y_0 \end{pmatrix} \begin{pmatrix} 4x_0^2cos(x_0^2) + 2sin(x_0^2) + y_0^2e^{x_0y_0} & x_0y_0e^{x_0y_0} + e^{x_0y_0} \\ x_0y_0e^{x_0y_0} + e^{x_0y_0} & x_0^2e^{x_0y_0} - 4 \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}$$

Subpart 3

Now to put x_0=0 and y_0=0 in the equation from subpart 2 (most of the expressions in the Hessian matrix and the gradient vector get equal to 0). So it becomes:

$$h(0,0) = -cos(0) + e^0 - 2 * 0$$

$$(\overrightarrow{x} - \overrightarrow{0})^T \overrightarrow{g}(0,0) = \begin{pmatrix} x-0 & y-0 \end{pmatrix} \begin{pmatrix} 0 * sin(0) + 0e^0 \\ 0 * e^0 - 0 \end{pmatrix}$$

$$\frac{1}{2}(\overrightarrow{x} - \overrightarrow{0})^T H(\overrightarrow{x} - \overrightarrow{0}) = \frac{1}{2} \begin{pmatrix} x-0 & y-0 \end{pmatrix} \begin{pmatrix} 0 * cos(0) + 2sin(0) + 0e^0 & 0 * e^0 + e^0 \\ 0 * e^0 + e^0 & 0 * e^0 - 4 \end{pmatrix} \begin{pmatrix} x-0 \\ y-0 \end{pmatrix}$$

$$0 + \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

if we do matrix multiplication now it gives:

$$\frac{1}{2}(xy + (x - 4y)(y))$$
$$\frac{1}{2}(xy + xy - 4y^2)$$
$$\frac{2xy - 4y^2}{2}$$
$$xy - 2y^2$$

Subpart 4

The Hessian from the last subpart at point (x_0, y_0) is:

$$\begin{pmatrix} 0 & 1 \\ 1 & -4 \end{pmatrix}$$

The two eigen values that we get from this Hessian Matrix are $\lambda = -\sqrt{5} - 2$ and $\lambda = \sqrt{5} - 2$. Which means that one of the eigen value is positive and the other one is negative, which makes the matrix **indefinite**.

Part B

Subpart 1

Taking the left hand side of the equation:

$$(\overrightarrow{x} - M^{-1}\overrightarrow{b})^T M(\overrightarrow{x} - M^{-1}\overrightarrow{b})$$

Taking the transpose inside the left paranthesis.

$$(\overrightarrow{x}^T - \overrightarrow{b}^T(M^{-1})^T)M(\overrightarrow{x} - M^{-1}\overrightarrow{b})$$

$(M^{-1})^T = M^{-1}$ because the matrix is symmetric, so its inverse is also symmetric.

$$(\overrightarrow{x}^T - \overrightarrow{b}^T M^{-1})M(\overrightarrow{x} - M^{-1}\overrightarrow{b})$$

Multiplying the middle M matrix to the left paranthesis.

$$(\overrightarrow{x}^T M - \overrightarrow{b}^T M^{-1}M)(\overrightarrow{x} - M^{-1}\overrightarrow{b})$$

$M^{-1}M = I$. Now multiplying the two paranthesis.

$$\overrightarrow{x}^T M\overrightarrow{x} - \overrightarrow{x}^T MM^{-1}\overrightarrow{b} - \overrightarrow{b}^T\overrightarrow{x} + \overrightarrow{b}M^{-1}\overrightarrow{b}$$

$$\overrightarrow{x}^T M\overrightarrow{x} - \overrightarrow{x}^T\overrightarrow{b} - \overrightarrow{b}^T\overrightarrow{x} + \overrightarrow{b}M^{-1}\overrightarrow{b}$$

Since both $\overrightarrow{x}$ and $\overrightarrow{b}$ are vectors their multiplication (with the first vector being transposed) gives us a 1×1 matrix (which is just multiplying all of the elements of both the vectors and adding the products - so it does not matter if x appears first or b).

$$(n \times 1)^T(n \times 1)$$
$$(1 \times n)(n \times 1)$$
$$(1 \times 1)$$

This tells us that both $-\overrightarrow{x}^T\overrightarrow{b}$ and $-\overrightarrow{b}^T\overrightarrow{x}$ are the same so adding them up.

$$\overrightarrow{x}^T M\overrightarrow{x} - 2\overrightarrow{b}^T\overrightarrow{x} + \overrightarrow{b}M^{-1}\overrightarrow{b}$$

Which is equal to the right hand side of the original equation.

Subpart 2

We have to prove that the following equation can be written as the simple quadratic equation:

$$f(\overrightarrow{x}) = (\overrightarrow{x} - \overrightarrow{\mu})^\top A(\overrightarrow{x} - \overrightarrow{\mu}) + (\overrightarrow{x} - \overrightarrow{\theta})^\top B(\overrightarrow{x} - \overrightarrow{\theta})$$

If I multiply the above equation as I did in the previous question, I will get the following:

$$f(x) = \|x\|A + \|x\|B + \|\mu\|A + \|\theta\|B - 2\overrightarrow{x}\overrightarrow{\mu}^T A - 2\overrightarrow{x}\overrightarrow{\theta}^T B$$

$$f(x) = -2\overrightarrow{x}\,\overrightarrow{\mu}^T A - 2\overrightarrow{x}\,\overrightarrow{\theta}^T B + [\|x\|A + \|x\|B + \|\mu\|A + \|\theta\|B]$$

The thing in the square brackets is already a number (constant) and as there are two vectors each in the first two terms, it can be said that the equation is written in the form of simple quadratic equation.

3 SVD and Eigendecomposition

a) Given a matrix $A = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$, find an analytical expression of A^n. Show your work. (You are allowed to use a computer to help you compute the eigenvalues/vectors for this part)

b) Consider the following matrix:

$$A = \begin{pmatrix} \frac{1+4\sqrt{3}}{4\sqrt{2}} & \frac{4-\sqrt{3}}{4\sqrt{2}} \\ \frac{4\sqrt{3}-1}{4\sqrt{2}} & \frac{4+\sqrt{3}}{4\sqrt{2}} \end{pmatrix}$$

(i) Show that the following is a **Singular Value Decomposition** of A:

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^\top$$

(ii) A matrix $R \in \mathbb{R}^{2\times 2}$ is a **2D rotation matrix** if it has the following form:

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where $\theta \in \mathbb{R}$

Geometrically speaking, $R_\theta \overrightarrow{v}$ rotates $\overrightarrow{v}$ counterclockwise by angle θ, for any $\overrightarrow{v} \in \mathbb{R}^2$, as shown in Figure 1.

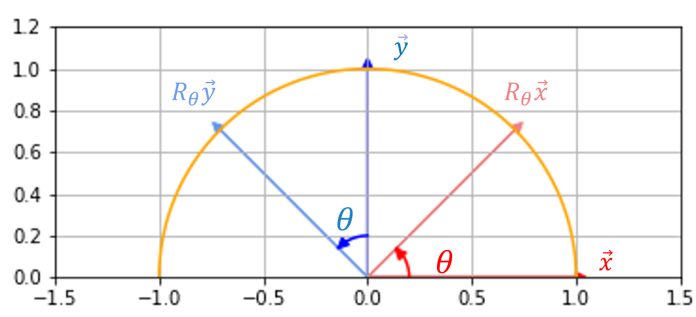


Figure 1: In this case, $\overrightarrow{x} = (1, 0)$ and $\overrightarrow{y} = (0, 1)$ are both rotated by $\theta = \frac{\pi}{4}$

Show that $U = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$ and $V^\top = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^\top$ are both rotation matrices, and find their corresponding rotational angles θ_U and θ_V.

(iii) Give an explanation for all geometric transformations performed by the SVD of A. In what order are the transformations performed?

(iv) The unit circle shown in Figure 2 has been transformed by a number of different 2D transformations. The transformation results are shown in Figure 3.

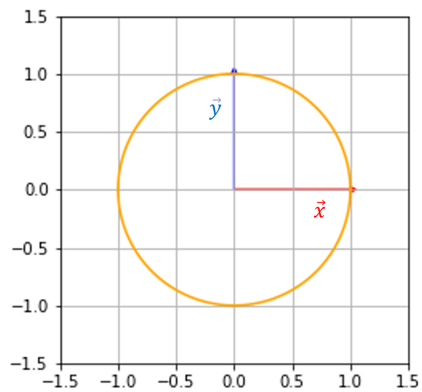
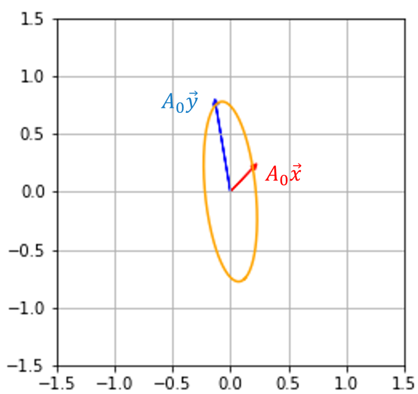
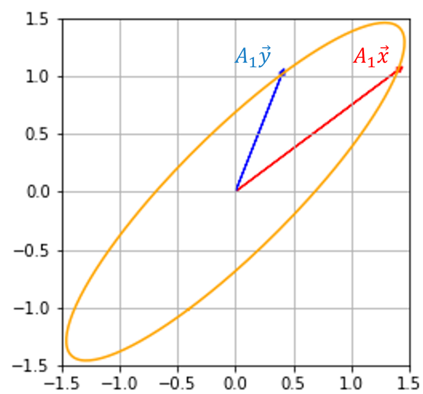


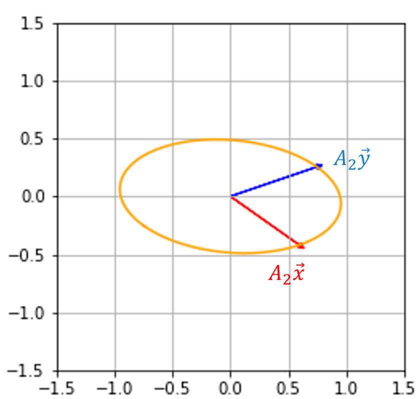
Figure 2: unit circle and plane basis $\overrightarrow{x}$, $\overrightarrow{y}$



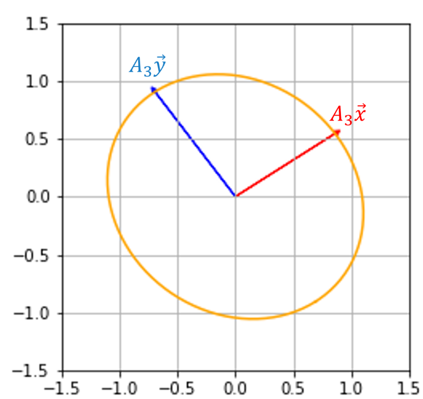
(a)



(b)



(c)



(d)

Figure 3: unit circle after different 2D transforms

Consider the geometric interpretation of the SVD for A. Which of the images from Figure 3 correctly shows the transformation of the unit circle by A? (i.e. Which of the matrices in the figure, A_1, A_2, A_3, A_4, is equal to A?) Please explain your choice.

Part A

As we know that if a matrix is multiplied by one of its eigenvectors ($\overrightarrow{x}$) it results in a number and the same eigenvector, essentially representing the original matrix with just a number, and that number is called the eigenvalue (λ) of the matrix.

$$A\overrightarrow{x} = \lambda\overrightarrow{x}$$

if we multiply A on both of the sides of the equation we get:

$$A^2\overrightarrow{x} = A\lambda\overrightarrow{x}$$

Now reordering the equation

$$A^2\overrightarrow{x} = \lambda A\overrightarrow{x}$$

Replacing Ax again with λx from the first equation

$$A^2\overrightarrow{x} = \lambda\lambda\overrightarrow{x}$$

$$A^2\overrightarrow{x} = \lambda^2\overrightarrow{x}$$

this can be extended to the power n.

$$A^n\overrightarrow{x} = \lambda^n\overrightarrow{x}$$

The interesting thing about this is that even if we increase the power of the A the eigenvectors of the matrix still stay the same and only the eigenvalues increase their power. Now to prove this I will consider the original matrix A and decompose it into its eigenvalues and eigenvectors and do the same for A^2 and A^n.

$$A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$$

its eigenvalues and eigenvectors are:

$$\lambda_1 = 5, \lambda_2 = 3 \quad \& \quad x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$A^2 = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

its eigenvalues and eigenvectors are:

$$\lambda_1 = 25, \lambda_2 = 9 \quad \& \quad x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$A^3 = \begin{pmatrix} 76 & 49 \\ 49 & 76 \end{pmatrix}$$

its eigenvalues and eigenvectors are:

$$\lambda_1 = 125, \lambda_2 = 27 \quad \& \quad x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

This proves that the follwing formula holds true for n power of the matrix. And the eigenvectors stay the same for a matrix with power n but the eigenvalues also take the power n.

$$A^n \overrightarrow{x} = \lambda^n \overrightarrow{x}$$

Part B

Subpart 1

$$A = \begin{pmatrix} \frac{1+4\sqrt{3}}{4\sqrt{2}} & \frac{4-\sqrt{3}}{4\sqrt{2}} \\ \frac{4\sqrt{3}-1}{4\sqrt{2}} & \frac{4+\sqrt{3}}{4\sqrt{2}} \end{pmatrix}$$

Working to find λ in $det(AA^T - \lambda I) = 0$.

Right multiplying the matrix A with its transpose.

$$AA^T = \begin{pmatrix} \frac{17}{8} & \frac{15}{8} \\ \frac{15}{8} & \frac{17}{8} \end{pmatrix}$$

Now doing $AA^T - \lambda I$.

$$AA^T - \lambda I = \begin{pmatrix} \frac{17}{8} - \lambda & \frac{15}{8} \\ \frac{15}{8} & \frac{17}{8} - \lambda \end{pmatrix}$$

the determinant is:

$$(\frac{17}{8} - \lambda)^2 - (\frac{15}{8})^2 = 0$$

$$\lambda^2 - \frac{17}{4}\lambda + 1 = 0$$

The values of σ_1 and σ_2 are: 4 & 1/4. And the Σ is the square root of the two σ s that we found.

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Now onto finding the two Orthogonal matrices U and V^T. The first of which is given by putting the singular values (not square root of the values) into the formula $AA^T - \sigma I$ (one by one). First putting $\sigma_1 = 1/4$.

$$\begin{pmatrix} \frac{17}{8} - \frac{1}{4} & \frac{15}{8} \\ \frac{15}{8} & \frac{17}{8} - \frac{1}{4} \end{pmatrix}$$

$$\begin{pmatrix} \frac{15}{8} & \frac{15}{8} \\ \frac{15}{8} & \frac{15}{8} \end{pmatrix}$$

Which makes our first column of matrix U be (the vectors are unit vectors because the matrix U is orthonormal):

$$u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Now putting the other value of $\sigma_2 = 4$.

$$\begin{pmatrix} \frac{17}{8} - 4 & \frac{15}{8} \\ \frac{15}{8} & \frac{17}{8} - 4 \end{pmatrix}$$

$$\begin{pmatrix} -\frac{15}{8} & \frac{15}{8} \\ \frac{15}{8} & -\frac{15}{8} \end{pmatrix}$$

Which makes our second column of matrix U be (the vectors are unit vectors because the matrix U is orthonormal):

$$u_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

The final U matrix is:

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Which can be simplified to:

$$U = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Now for the last orthogonal matrix. The following formula should be used:

$$A^T = (U\Sigma V)^T$$
$$A^T = V^T \Sigma^T U^T$$

For orthogonal matrices $U^T = U^{-1}$. So moving U^T to the other side and it gives us:

$$A^T U = V^T \Sigma^T$$

$\Sigma^T = \Sigma$ for a diagonal matrix.

$$A^T U = V^T \Sigma$$

$$A^T U = \begin{pmatrix} \sqrt{3} & -\frac{1}{4} \\ 1 & \frac{\sqrt{3}}{4} \end{pmatrix}$$

This matrix gives us the matrix V multiplied by Σ. Which if we separate out gives us:

$$\begin{pmatrix} \sqrt{3} & -\frac{1}{4} \\ 1 & \frac{\sqrt{3}}{4} \end{pmatrix} => \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Which makes the matrix V equal to:

$$V = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

So, the equation asked in the question checks out.

$$A = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^T$$

Subpart 2

To show that the both matrices U and V^T are rotation matrices, I will map each individual values of the matrix to the actual rotation matrix. First I will go with the Matrix U.

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

If we put $\theta = 45^o$ we get:

$$cos(45^o) = \frac{\sqrt{2}}{2}$$
$$-sin(45^o) = -\frac{\sqrt{2}}{2}$$
$$sin(45^o) = \frac{\sqrt{2}}{2}$$
$$cos(45^o) = \frac{\sqrt{2}}{2}$$

Which is exactly what is written in the matrix U. So the matrix U is a rotation matrix and it rotates by an angle of 45^o.

Now to check if the matrix V^T has the same output.

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^T$$

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

If we put $\theta = -30^o$ we get:

$$cos(-30^o) = \frac{\sqrt{3}}{2}$$
$$-sin(-30^o) = \frac{1}{2}$$
$$sin(-30^o) = -\frac{1}{2}$$
$$cos(-30^o) = \frac{\sqrt{3}}{2}$$

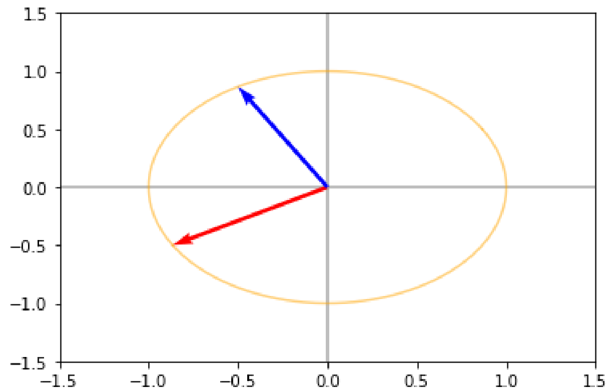
So we can conclude that the matrix U is a rotation matrix (which rotates by an angle of 45^o) and V^T is also a rotation matrix (which rotates by an angle of -30^o).

Subpart 3

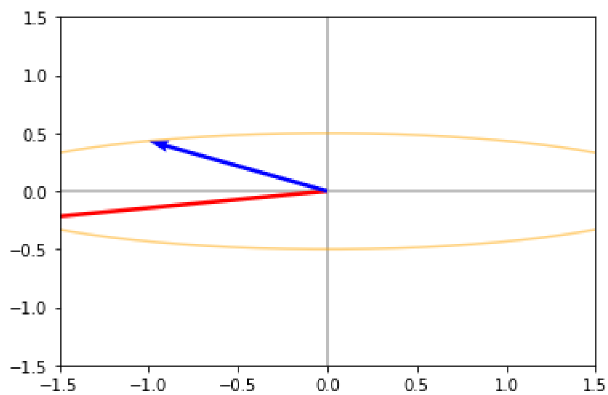
The order of the transformation is rotation, scaling and then again rotation. SVD applies rotation of 45^o to the circle which makes it move in the counterclockwise direction. After that the Σ scales the matrix so that the circle expands twice in one direction (x) and shrinks twice in the other direction (y) because the singular values of the Σ are 2 and 1/2. After the first two operations are done, the now oval shape is rotated 30^o in the clockwise direction (opposite to the previous rotation because of the negative sign [-30^o]). Images for these transformations are provided in the Subpart 4.

Subpart 4

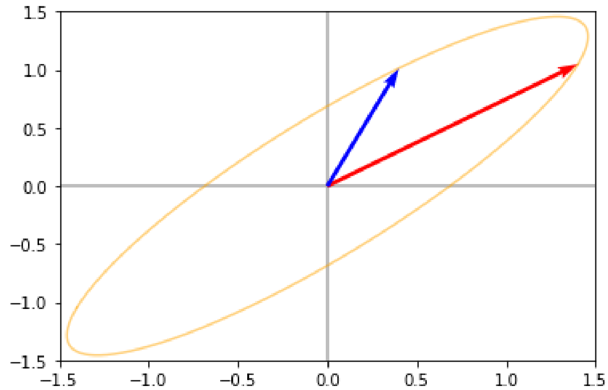
The matrix from the options that represents the original matrix A is A_2 or the option b. As the order of the operations is rotation scaling and then again rotation. The first rotation causes the unit circle to rotate in the counter clockwise direction. After which the circle looks something like this:



After that the matrix is then scaled by the order of 2 and 1/2. The 2 in the Σ makes the circle grow twice as big as it was before in the x direction and the 1/2 makes the circle shrink down to half of its original size. After scaling the circle looks something like this:



After the first two operations the circle is rotated in the opposite direction (as proved in the Subpart 2). And the final circle looks something like this:


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5