Jae Hong lee
Prof. Tsourakakis
Boston University

April 16, 2024

— **CAS CS 365 - Problem Set 2**

Collaborators: Wyatt Naiper, Aditya Chowdhri

1 SVD [25 Points]

1.1 Problem 1.1

Find the SVD of A=[1, 1] without the use of computing devices/software

A is 1×2 matrix. Since the $Rank(A) = 1$, The dimensions of U, Σ, V^T are

$U = 1 \times 1$, $\Sigma = 1 \times 2$, $V^T = 2 \times 2$. Since A is composed with two 1 vectors. $U = \begin{bmatrix} 1 \end{bmatrix}$.

Σ is the diagonal matrix so the second column of Σ is 0, and $\Sigma = \begin{bmatrix} 1 & 0 \end{bmatrix}$.

In order to make $A = \begin{bmatrix} 1, & 1 \end{bmatrix}$, $V^T = \begin{bmatrix} 1, & 1 \end{bmatrix}$. Then $V = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Normalize it to $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. This makes Σ to $\Sigma = \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix}$.

V is orthonormal, and 2×2 matrix, so add one more column $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ to make it orthonormalized. Now $V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

$A = \begin{bmatrix} 1 \end{bmatrix} \times \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix} \times \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

1.2 Problem 1.2

Let $A \in R^{m \times n}$ let σ_1 be the maximum singular value of A. For $x \in R^n \setminus \{0\}$ the spectral norm of A is defined as $||A||_2 = max_x \frac{||Ax||_2}{||x||_2}$. Prove that

$$||A||_2 = \sigma_1$$

In order to show $||A||_2 = \sigma_1$, I need to show two parts:

$$(i)||A||_2 \leq \sigma_1$$

$$(ii)||A||_2 \geq \sigma_1$$

Part (i): $||A||_2 \leq \sigma_1$

$\frac{||Ax||_2}{||x||_2} = \frac{\sqrt{x^T A^T A x}}{\sqrt{x^T x}}$. $A = U \times \Sigma \times V^T$ so $A^T A = V \times \Sigma^2 \times V^T$. so $\frac{x^T A^T A x}{x^T x} = \frac{x^T \times V \times \Sigma^2 \times V^T \times x}{x^T x} = \frac{||\Sigma V^T x||_2}{||x||_2}$.

Let say $y = V^T x$, $||y||^2 = y^T y = x^T V \times V^T x = x^T x = ||x||^2$. So even if $y = V^T x$, $||y||_2 = ||x||_2$.

we can apply this to $\frac{||\Sigma V^T x||_2}{||x||_2}$.

$\frac{||\Sigma V^T x||_2}{||x||_2} = \frac{||\Sigma y||_2}{||y||_2}$. Since Σ is a diagonal matrix, its effect is only to scale the vector y along its axes.

$y = [y_1, y_2, y_3, ..., y_n]^T$.

And $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \vdots & \dots & \sigma_n \end{bmatrix}$, $\Sigma^2 = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \vdots & \dots & \sigma_n^2 \end{bmatrix}$.

$\sigma_1, \sigma_2, ..., \sigma_n$ are the singular values of A in non-decreasing order. Since σ_1 is the maximum singular value, σ_1^2 is the maximum term.

So $\frac{||\Sigma y||_2}{||y||_2} = \frac{\sqrt{\sum_{i=1}^n \sigma_i^2 y_i^2}}{\sqrt{\sum_{i=1}^n y_i^2}} \leq \frac{\sqrt{\sigma_1 \sum_{i=1}^n y_i^2}}{\sqrt{\sum_{i=1}^n y_i^2}} = \sqrt{\sigma_1^2} = \sigma_1$.

Part (ii) : $||A||_2 \geq \sigma_1$

Let say the vector $x = V \times i_1$, i_1 is the first canonical basis vector in R^n. Then $||x||_2 = ||Vi_1||_2 = ||i_1||_2 = 1$.

Then for $||Ax||_2$,

$$||Ax||_2 = ||U\Sigma V^T V i_1||_2 = ||U\sigma_1 i_1||_2 = \sigma_1$$

So at particular vector x, $||Ax||_2 = \sigma_1$. As σ_1 is the maximum singular value, $||Ax||_2 \geq \sigma 1$.

Combining Part(i) and Part(ii), $||A||_2 = \sigma_1$

1.3 Problem 1.3

Suppose you are given a system of linear equations $A^{m\times n}x^{n\times 1} = b^{m\times 1}$ where the number of rows m is greater than the number of columns n (overdetermined system of linear equations). Given that the number of equations m is greater than the number of unknowns maybe there is no x that satisfies the linear system. Thus it is natural to try to find an x that minimizes the error $||Ax - b||_2$.

1.3.1 Problem 1.3a

Assume that A is full rank, i.e., $rank(A) = n < m$. Prove that the unique minimizer $x^* = (A^TA)^{-1}A^Tb$. Be explicit about where you use the assumption that A is full rank and write down the optimal objective value $||Ax^* - b||_2$.

$\|Ax - b\|_2 = x^TA^T \times Ax - 2x^TA^Tb + \|b\|^2$. We need to try to find an x that minimize the error, so derivative of $\|Ax - b\|_2$ to be 0.

$$\begin{aligned} (\|Ax - b\|_2)' = (x^TA^T \times Ax - 2x^TA^Tb + \|b\|^2)' &= 2A^TAx - 2A^Tb \\ 2A^TAx - 2A^Tb &= 0 \\ 2A^TAx &= 2A^Tb \\ A^TAx &= A^Tb \end{aligned}$$

A is a full rank $m \times n$ matrix. A is a full rank matrix, so A^TA is a full rank. Since it is the full rank matrix, AA^T is the invertible.

$$\begin{aligned} A^TAx &= A^Tb \\ (A^TA)^{-1}A^TAx &= (A^TA)^{-1}A^Tb \\ x &= (A^TA)^{-1}A^Tb \end{aligned}$$

I have showed the unique minimizer $x^* = (A^TA)^{-1}A^Tb$.

1.3.2 Problem 1.3b

Use the SVD decomposition to solve the same optimization problem when A is rank deficient.

When A is rank deficient, meaning it has fewer linearly independent rows than columns.

Let $A = U\Sigma V^T$ be the singular value decomposition of A, where U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ diagonal matrix with singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ (where r is the rank of A) arranged in descending order along its diagonal, and V is an $n \times n$ orthogonal matrix.

Since A is rank deficient, some singular values in Σ will be zero.

$||Ax - b||_2 = ||(U \times \Sigma \times V^T)x - b||_2$

Let's say there is a matrix $U^T y$, for $y \in R^M$. Then $||U^T y||_2 = \sqrt{(y^T U) \times (U^T y)} = \sqrt{y^T \times y} = ||y||)_2$. the norm of y and $U^T y$ are both equal to $||y||_2$.

Let apply this here

$$
\begin{aligned}
||Ax - b||_2 &= ||U^T(AVV^T x - b)||_2 && (V \times V^T = I) \\
&= ||\Sigma(V^T x) - U^T b||_2 \\
&= ||\Sigma z - U^T b||_2 && z = V^T x \\
&= \sqrt{\sum_{i=1}^{r} (\sigma_i z_i - u_i^T b)^2 + \sum_{j=r+1}^{m} (u_j^T b)^2}
\end{aligned}
$$

The optimal solution is given by $z_i = \frac{u_i^T b}{\sigma_i}$, $i = 1, ..., r$. And the object becomes $\sqrt{\sum_{i=r+1}^{m} (u_i^T b)^2} = min_x ||Ax - b||_2$.

The actual $x^* = Vz^* = x^* = \sum_{i=1}^{r} (\frac{u_i^T b}{\sigma_i}) v_i$

1.4 Problem 1.4

What is the SVD of the matrix $M = [0, 1, 2]^{1\times 3}$? Calculate it in three different ways: (i) by directly observing the matrix, (ii) by using MM^T, and (iii) by using $M^T M$.

1.4.1 (i) by directly observing the matrix

The matrix M is 1×3 matrix. The rank of M is 1 since there is only one row. So the dimensions of U, Σ, V^T are $U = 1 \times 1$, $\Sigma = 1 \times 1$, $V^T = 1 \times 3$.

U is $\begin{bmatrix} 1 \end{bmatrix}$ since U is 1×1 vector.

$\Sigma \times V^T = \begin{bmatrix} 0, & 1, & 2 \end{bmatrix}$, and V^T has to be orthonormal, the length of row of V^T (which is the **column** length of vector V) has to be 1. So normalizead row is $\begin{bmatrix} 0, & \frac{1}{\sqrt{5}}, & \frac{2}{\sqrt{5}}, \end{bmatrix}$.

Now Σ is $\sqrt{5}$ in order to make the product of $U \times \Sigma \times V^T = \begin{bmatrix} 0, & 1, & 2 \end{bmatrix}$.

$\therefore U = \begin{bmatrix} 1 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sqrt{5} \end{bmatrix}$, $V^T = \begin{bmatrix} 0, & \frac{1}{\sqrt{5}}, & \frac{2}{\sqrt{5}}, \end{bmatrix}$

1.4.2 (ii) by using MM^T

The matrix M is 1×3 matrix. M^T is 3×1 vector. MM^T is 1×1 vector.

$M = \begin{bmatrix} 0, & 1, & 2 \end{bmatrix}$, $M^T = \begin{bmatrix} 0, \\ 1, \\ 2 \end{bmatrix}$. $M \times M^T = \begin{bmatrix} 5 \end{bmatrix}$

Also $M = U \times \Sigma \times V^T$, $M^T = V \times \Sigma \times U^T$.

$M \times M^T = U \times \Sigma \times V^T \times V \times \Sigma \times U^T = U \times \Sigma^2 \times U^T$.

In previous question, we know the dimensions of U and Σ are both $U = 1 \times 1$, $\Sigma = 1 \times 1$.

We know U need to be orthnormal, So $U = \begin{bmatrix}1\end{bmatrix}$.

$MM^T = U \times \Sigma^2 \times U^T = 1 \times \Sigma^2 \times 1 = 5$, so $\Sigma = \begin{bmatrix}\sqrt{5}\end{bmatrix}$.

Since we know $U = \begin{bmatrix}1\end{bmatrix}$ and $\Sigma = \begin{bmatrix}\sqrt{5}\end{bmatrix}$, V^T has to be $\begin{bmatrix}0, & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{bmatrix}$ to make $M = \begin{bmatrix}0, & 1, & 2\end{bmatrix}$

$\therefore U = \begin{bmatrix}1\end{bmatrix}$, $\Sigma = \begin{bmatrix}\sqrt{5}\end{bmatrix}$, $V^T = \begin{bmatrix}0, & \frac{1}{\sqrt{5}}, & \frac{2}{\sqrt{5}},\end{bmatrix}$

1.4.3 (iii) by using $M^T M$

The matrix M is 1×3 matrix. M^T is 3×1 vector. $M^T M$ is 3×3 matrix.

$M = \begin{bmatrix}0, & 1, & 2\end{bmatrix}$, $M^T = \begin{bmatrix}0, \\ 1, \\ 2\end{bmatrix}$. $M^T \times M = \begin{bmatrix}0, & 0, & 0 \\ 0, & 1, & 2 \\ 0, & 2, & 4\end{bmatrix}$

$M^T M = V \times \Sigma \times U^T \times U \times \Sigma \times V^T = V \times \Sigma^2 \times V^T$.

In previous question 1.4 (i), we know the dimensions of V and Σ are $V = 3 \times 1$, $\Sigma = 1 \times 1$.

Let say $V = \begin{bmatrix}v1 \\ v_2 \\ v_3\end{bmatrix}$. Extends the Equation $V \times \Sigma^2 \times V^T = \begin{bmatrix}v1 \\ v_2 \\ v_3\end{bmatrix} \times \begin{bmatrix}\Sigma\end{bmatrix} \times \begin{bmatrix}\Sigma\end{bmatrix} \times \begin{bmatrix}v_1, & v_2, & v_3\end{bmatrix} =$ $\begin{bmatrix}v_1\Sigma\Sigma v_1 & v_1\Sigma\Sigma v_2 & v_1\Sigma\Sigma v_3 \\ v_2\Sigma\Sigma v_1 & v_2\Sigma\Sigma v_2 & v_2\Sigma\Sigma v_3 \\ v_3\Sigma\Sigma v_1 & v_3\Sigma\Sigma v_2 & v_3\Sigma\Sigma v_3\end{bmatrix} = \begin{bmatrix}0, & 0, & 0 \\ 0, & 1, & 2 \\ 0, & 2, & 4\end{bmatrix}$

From the extension, we know $v_1 = 0, v_2 = 1, v_3 = 2$ and $\Sigma = \begin{bmatrix}1\end{bmatrix}$. The length of v_1, v_2, v_3 need to be 1, so normalized V is $\begin{bmatrix}0, \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}}\end{bmatrix}$ and Σ is $\begin{bmatrix}\sqrt{5}\end{bmatrix}$.

U has to be $\begin{bmatrix}1\end{bmatrix}$ to make $M = \begin{bmatrix}0, & 1, & 2\end{bmatrix}$

$\therefore U = \begin{bmatrix}1\end{bmatrix}$, $\Sigma = \begin{bmatrix}\sqrt{5}\end{bmatrix}$, $V^T = \begin{bmatrix}0, & \frac{1}{\sqrt{5}}, & \frac{2}{\sqrt{5}},\end{bmatrix}$.

2 Reservoir Sampling [10 points]

Design an algorithm that samples $k \geq 1$ elements uniformly at random from an insert-only stream, whose length is unknown. Present the pseudocode and prove the correctness of the proposed algorithm.

Algorithm 1 Reservoir Sampling

```
Input : Stream, k
Array array[size k]
int i ← 0
while stream[i] == True do
    if i ≤ k then
        array.append Stream[i]
    else
        if k/i ≤ Uniform[0,1] then
            j ← rand(k)
            array[j] = stream[i]
        end if
    end if
    i++
end while
Return Array
```

In my pseudocode, the algorithm reads down the input stream, and $\frac{k}{i}$ probability of chance to get a new element and replace the one of previous element with the array uniform chance $\frac{1}{k}$ and $\frac{i-k}{i}$ chance of probability to keep the previous array and next to the next stream.

Prove

Let's say there is a stream $A = A_1, A_2, ..., A_{unknown}$. ith sampling result is S_i.

Followed by my algorithm, the probability of $strean[i]$ element in the array with ith sampling is $\frac{k}{n}$.

Then the probability of $stream[i-1]$ element in the array with ith sampling is two parts

$$\begin{cases} \text{choosing } i\text{th sampling but } i-1 \text{ element is not replaced: } \frac{k}{i} \times \frac{k-1}{k} = \frac{k-1}{i} \\ \text{not choosing } i\text{th sampling so keep } i-1 \text{ element in the array: } \frac{i-k}{i} \end{cases}$$

The sum of two probability is $\frac{k-1}{i} + \frac{i-k}{i} = \frac{i-1}{i}$. However, these both probabilities has the assumption that $i-1$ element is in the array. Which is the probability of $\frac{k}{i-1}$. So the probability of $stream[i-1]$ element in the array with ith sampling is $\frac{k}{i-1} \times \frac{i-1}{i} = \frac{k}{i}$.

The probability calculation is applied to all element each first k elements so the probability is as follows:

Let's say the probability of each element in the k length array is S_i.

$$S_i = \begin{cases} \frac{k}{i} \\ 1 \text{ if } i \leq k \end{cases}$$

3 Challenges with Coding [65 points]

The coding part of the homework is available on Git here. Follow the instructions in the Jupyter notebook. After finishing it, download the notebook in PDF format, which should include all the code and figures. If you have trouble converting it to PDF, you can download it as HTML, and then save it as PDF. Submit the math problems and coding problems separately in two files through Gradescope.

Submitted by Jae Hong lee on April 16, 2024.