Topic 3: Singular Value Decomposition (SVD)

Let $A \in \mathbb{R}^{m \times n}$ be any matrix. The SVD allows to interpret (break down) the action of A on vectors (i.e as a linear operation) as a sequence of three simpler steps. i.e.) as action of 3 different structured matrices:

1. Reflection / Rotation (orthogonal matrix) in the domain $\mathbb{R}^n$
2. Scaling (diagonal operator)
3. Reflection / Rotation (orthogonal matrix) in the image space $\mathbb{R}^m$

This is similar to the eigenvalue decomposition by applying the following tradeoffs:

In the eigenvalue decomposition (EVD), we write A as VDV^{-1}, where A is square, V, D have the same dimensions as A, and V is invertible. Here V is a change of basis matrix allowing to map the operator to a diagonal version of it.

i.e) map vectors to standard basis vector, then apply the scaling from D, and the nrevert the change of basis operation.

The EVD has a numer of drawbacks:

1. Does not always exist. e.g) for non-square matrices
2. Does not alays exist even for square matrices. e.g) missing eigenvectors for eigenvalues of multiplicity > 1
3. You need complex geometry.

In the **SVD**

- **Give up:** the same matrix V on both sides
- **Gain:**
    a. Always exist for **any** matrix square & non-square
    b. If A is real, so is the SVD. So the geometry is expressible in real terms.
    c. The diagonal part of the SVD has non-negative elements, which you can choose to be ordered.
    d. **Singular vectors always exist.**

In some special cases, the SVD & EVD are the same.

We first state the SVD as a theorem, explore some of its implications & interperations, understand its geometry, & and then prove it, & also see an algorithm for computing it.

Theorem: Existence of SVD

Let $A \in \mathbb{R}^{m \times n}$ be any matrix.

There exists:

1. $U \in \mathbb{R}^{m \times m}$ orthogonal
2. $\Sigma = diag(\sigma_1, \sigma_2, ... \sigma_{min(m,n)}) \in \mathbb{R}^{m \times n}$, where:
    - $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{min(m,n)} \geq 0$
3. $V \in \mathbb{R}^{n \times n}$ orthogonal,

    s.t $A = U \Sigma V^T$

    - Note: here, V^T is a notational convenience

There are several ways of expressing this relationship. i.e) several ways to write SVD.

1. $A = U \Sigma V^T$, is called the **full SVD**
2. **Reduced SVD:**
    a. **Case 1:** $m \geq n$ (case tall)

        Take:

        $U = [U_1, U_2]$, U_1 has shape $m \times n$, U_2 has shape $m \times (m - n)$ -> m x m eventually

        $\Sigma = \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}$, Σ_1 has shape $n \times n$, 0 has shape $(m - n) \times n$ -> m x n eventually

        $V \rightarrow$ no need to break down, n x n.

        Then (m x m) (m x n) (n x n) -> (m x n), which is same shape as A:

$$
\begin{aligned}
A &= U \Sigma V^T \\
&= [U_1, U_2] \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V^T \\
&= (U_1 \Sigma_1 + U_2 0) V^T \\
&= U_1 \Sigma_1 V^T
\end{aligned}
$$

    b. **Case 2:** $m < n$ (case wide)

        $U \rightarrow$ leave unchanged

$\Sigma = [\Sigma_1, 0]$, Σ_1 has shape $m \times n$, 0 has shape $m \times (m - n)$

$V = [V_1, V_2]$, V_1 has shape $n \times m$, V_2 has shape $n \times (n - m)$

We then get (m x m) (m x n) (n x n) -> (m x n):

$$\begin{aligned} A &= U\Sigma V^T \\ &= U[\Sigma_1 \ 0] \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\ &= U\Sigma_1 V_1^T \end{aligned}$$

c. **Case 3:** SVD as a sum of rank-1 matrices

$U = [u_1, u_2 \ldots . u_m], u_i \in \mathbb{R}^m \to (m \times m)$

$\Sigma = diag(\sigma_1, \sigma_2, \ldots \sigma_{min(m,n)}), \Sigma \in \mathbb{R}^{m \times n}$ (Unchanged)

$V = [v_1, v_2, \ldots v_n], v_j \in \mathbb{R}^n \to (n \times n)$

Then we have:

$$\begin{aligned} A &= U\Sigma V^T \\ &= [u_1, u_2 \ldots . u_m] diag(\sigma_1, \sigma_2, \ldots \sigma_{min(m,n)}) \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} \\ &= [u_1, u_2 \ldots . u_m] \begin{bmatrix} \sigma_1 v_1^T \\ \sigma_2 v_2^T \\ \vdots \\ \sigma_{min(m,n)} v_{min(m,n)}^T \\ 0 \\ 0 \end{bmatrix} \text{(possibly the zero terms)} \\ &= \sum_{k=1}^{min(m,n)} \sigma_k u_k v_k^T \end{aligned}$$

- **Note:** each $\sigma_k, k \in [1, min(m, n)]$ is a real value, $u_k v_k^T$ is a shape (m x 1) (1 x n) matrix multiplication, which results in a (m x n) matrix. Thus the sum equals to a (m x n) matrix scalred with all σ_k values.
- **Example** of $\Sigma = diag(\sigma_1, \ldots \sigma_{min(m,n)})$

  Say we have A being 3 by 2, then even Σ is 3 by 2 gurantee, but we only have 2 values: σ_1, σ_2

  and the whole Σ will look like:

$$\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}$$

Definitions of SVD components

1. $\sigma_1, \sigma_2, ... \sigma_{min(m,n)}$ are called the **<u>singular values</u>** of matrix A
2. $v_1, ... v_n$ are called the **<u>right singular vectors</u>** of A
3. $u_1, ... u_m$ are called the **<u>left singular vectors</u>** of A

Note:

1. The SVD can also be defined for complex matrices. i.e) $A \in \mathbb{C}^{m \times n}$ can be written as $A = U\Sigma V^H$ with u, v `unitary`. Σ diagonal real $\Longrightarrow$ same form as before.
    - H : complex conjugate transpose.
    - **<u>unitary:</u>** $u^H u = uu^H = I_m, v^H v = vv^H = I_n$
2. **<u>Is SVD unique????</u>**

    Generally, NO!

    Gennerally, SVD is <u>not</u> unique. e.g) in the *reduced* vs *full* expression we have zero blocks (e.g. tall case u_2 is an arbitrary completion, starting from u_1 to the orthogonal matrix U)

    Also, e.g. in the **rank-1 expression,** $\sigma_k u_k v_k^T = \sigma_k(-u_k)(-v_k)^T$

    However!

    $\sigma_1, \sigma_2, ... \sigma_{min(m,n)}$, the **singular values are unique** (with the constraints **<u>positive, decreasing</u>**).

    As well, the left and right singular vectors corresponding to **<u>non-zero</u>** singular values are **unique depending on sign (+/-) choices.**

Example of Uses of SVD

Set up

Write the rank-1 representation, but remove any zero singular values, so, that is to say:

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

Where:

1. $r \leq min(m, n)$
2. $\sigma_1 \geq \sigma_2 \geq .... \geq \sigma_r > 0$

Ex 1: Image of A (as a linear transformation)

> *We take the $v_1, ... v_n$, the right singular vectors and express any vector in $\mathbb{R}^n$ by a linear combination with them since we know that SVD provides a orthgonal matrix $V \implies$ orthonormal column space $\implies$ orthonormal basis of column space.*

Let $x = \alpha_1 v_1 + ... + \alpha_n v_n \in \mathbb{R}^n$, here we have ($v_1, ... v_n$ are an orthonormal set & form a basis for $\mathbb{R}^n$)

$$\begin{aligned} Ax &= \sum_{k=1}^{r} \sigma_k u_k v_k^T (\alpha_1 v_1 + ... + \alpha_n v_n) \\ &= \sum_{k=1}^{r} \sigma_k \alpha_k u_k \end{aligned}$$

- since $v_k^T v_j = 0$ if $j \neq k$ and $v_k^T v_j = 1$ if $j = k$

Here as $\alpha_k' s$ range over $\mathbb{R}$, $\sigma_k \alpha_k$ also range over $\mathbb{R}$, thus $Image(A) = span\
{u_1,u_2,...u_r\}$

- where $u_1, u_2, ... u_r$ are from the left singular vectors

<u>Rank:</u>

$col\ rank(A) = dim\ Image(A) = r = \#$ of non-zero singular values

Here $row\ rank(A) = r = col\ rank(A)$ because the singular values for A^T are the same as for A since $A^T = V\Sigma^T U^T$

- Identical $\sigma_k' s$ singular values but arranged in a slightly different way

Geometry of Action of A using SVD

To understand the geometry, first look at the action on vectors from a diagonal matrix Σ. Say, as an example:

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, ||x||_2 = 1$$

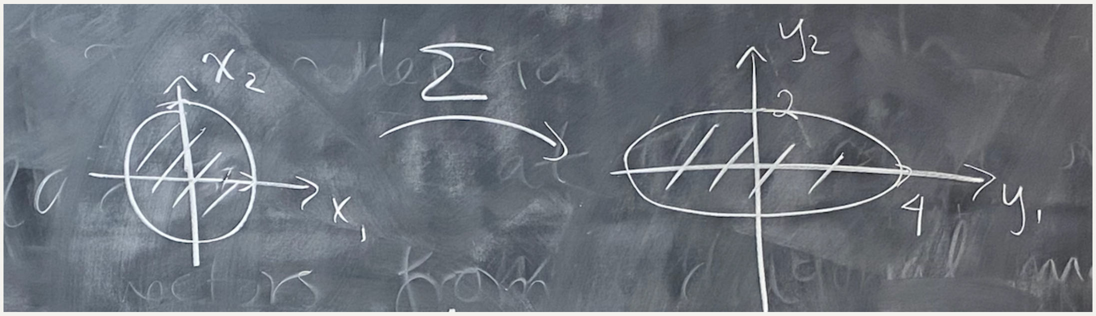
Then we have:

$$\Sigma x = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y$$

and it maps the unit circle to another circle of radius 2 (y).

It also maps $||x||^2 \leq 1$ to vectors y s.t $||y||^2 \leq 4$. i.e) circle to circle & full ball to full ball (i.e interiro)

What about $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$, then we have $y = \begin{pmatrix} 4x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$



This maps circle to ellipse both boundary & interior.

This is an ellipse because $1 = ||x||^2 = x_1^2 + x_2^2 = (\frac{y_1}{4})^2 + (\frac{y_2}{2})^2$.

- The boundary here is captured by the relation with **equality**
- & the interior is captured if you use $1 \geq ||x||^2 = ||x||^2 = x_1^2 + x_2^2 = (\frac{y_1}{4})^2 + (\frac{y_2}{2})^2$
  - *interior:* refers to all the points inside and on the boundary of the shape

Another example with a 0 diagonal element:

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \text{ take } 1 = ||x||^2 = x_1^2 + x_2^2 + x_3^2$$

then we have:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \Sigma x = \begin{pmatrix} 3x_1 \\ 2x_2 \\ 0 \end{pmatrix}$$

Here, we thus have:

$$\begin{aligned} 1 &= x_1^2 + x_2^2 + x_3^2 \\ &\geq x_1^2 + x_2^2 \\ &= (\frac{y_1}{3})^2 + (\frac{y_2}{2})^2 \text{ boundary} \\ OR \\ &\geq (\frac{y_1}{3})^2 + (\frac{y_2}{2})^2 \text{ interior} \end{aligned}$$

When "reducing" dimension, you may map boundary (3 dimensional, since x_1, x_2, x_3) to boundary (2 dimensional with equal sign, y_1, y_2) or interior with bigger or equal to (y_1, y_2).

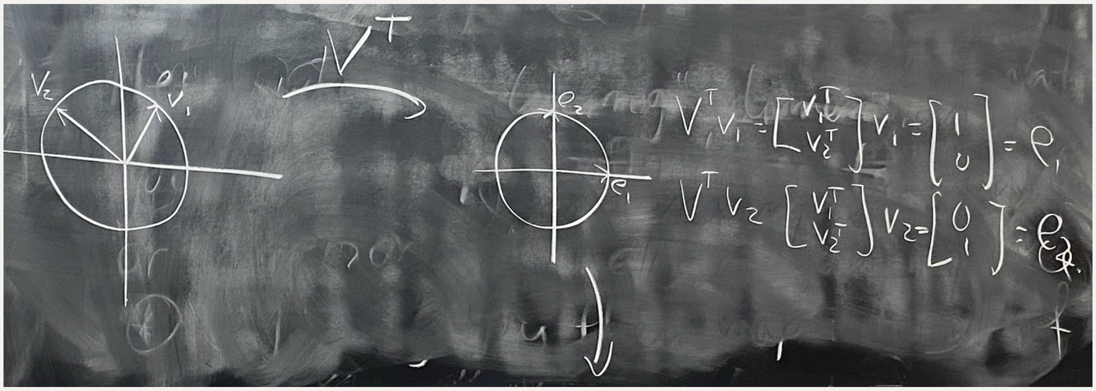
As well if you replace the equal sign by:

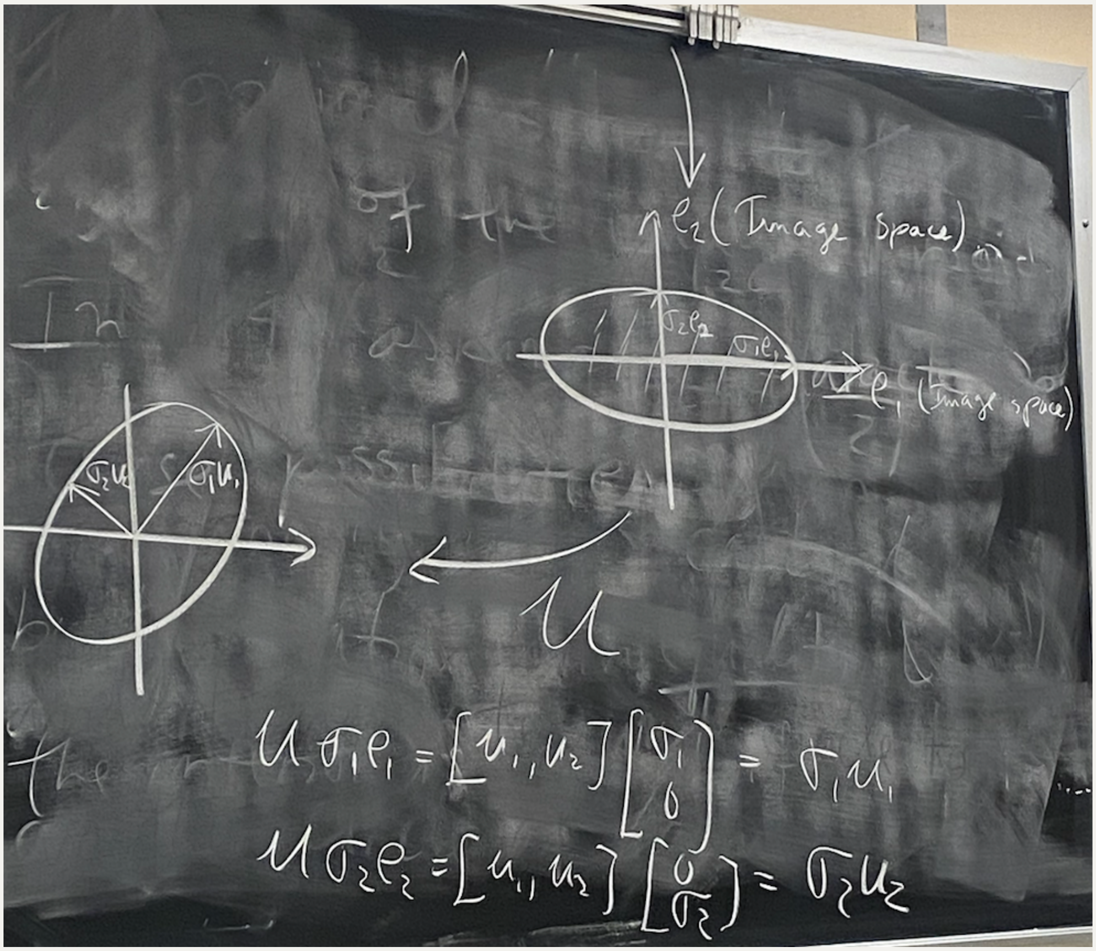
$1 \geq \ldots$, you then map interior of the original sphere to the interior of the image ellipsoid. In $A4$, aksing to characterize these possibilities.

$$\begin{aligned} 1 &\color{red}{\geq} x_1^2 + x_2^2 + x_3^2 \text{ interior} \\ &\geq x_1^2 + x_2^2 \\ &\geq (\frac{y_1}{3})^2 + (\frac{y_2}{2})^2 \text{ interior} \end{aligned}$$

Now adding the other components of the SVD (the left/right singular vectors) only **reorients the shapes**, first in the domain through V, and later in the image-space through U.

Putting all together, the action of a matrix produces the following geometrical schema:





Matrix Norms (prerequisite for proof of SVD)

The $m \times m$ matrices with addition and scalar multiplication can be packaged into a vector space, so that you can define norms on them.

There are **two ways** to do so:

1. You can just define arbitrary norms, e.g. **Frobenius norm (F-norm)**

- **Def:** Given $A \in \mathbb{R}^{m\times n}$, the Frobenius norm is the following: (sum up all entries's square, then square root them all)

$$||A||_F = \left(\sum_{\substack{i=1,..m \\ j=1,...n}} a_{ij}^2 \right)^{\frac{1}{2}}$$

2. **Induced norm**, (i.e. induced by an underlying vector)

   - **Def:** Let || || be a vector norm (1-norms, 2-norms) on $\mathbb{R}^n$, $\mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$. The induced matrix norm on A is defined by:

   $$||A|| = \max_{\substack{x\neq 0 \\ x\in\mathbb{R}^n}} \frac{||Ax||}{||x||} = \max_{\substack{||x||=1 \\ x\in\mathbb{R}^n}} ||Ax|| = \text{ maximum stretch (scaling) that A can appl}$$

   Here, || || notation is overloaded becase $||x||$ applies to $\mathbb{R}^n$ and $||Ax||$ applies to $\mathbb{R}^m$, and $||A||$ has a dependency on each on above.

   Usually people keep to one "kind" of norm in $\mathbb{R}^m \& \mathbb{R}^n$, e.g. 2-norm (which means, we do not change how to compute the norm of vectors when changing between ranges)

3. ***Side Question:***

   - Is the Frobenius norm induced? How do we know?

   The F-norm is not induce because:

   e.g) Let's compute the F-norm and induced norm of identity matrix:

   i. $$||I_m||_F = \sqrt{(\text{ all elements })} = \sqrt{m}$$

   ii. $$||I_m|| = \max_{||x||=1} ||Ix|| = \max_{||x||=1} ||x|| = 1$$

   Therefore the F-norm can not be induced.

More Preparation for proof of SVD

Lemma 1

back to proof

Let $A \in \mathbb{R}^{m\times n}$

Then we have:

$$||A||_2 \overset{1}{=} \max_{\substack{x\in\mathbb{R}^n \\ y\in\mathbb{R}^m \\ ||x||_2=||y||_2=1}} y^TAx \overset{2}{=} \max_{\substack{x\in\mathbb{R}^n \\ y\in\mathbb{R}^m \\ ||x||_2=||y||_2=1}} x^TA^Ty \overset{3}{=} ||A^T||_2$$

- ***where both** $||A||_2, ||A^T||_2$ **are induced***
- **Note:**
    - Within the proof (see below), we found (by existence argument) a v, such that with

      $||v||_2 = 1, x = v, y = \frac{Av}{||Av||_2}$, we achieve these bounds.
    - This v was defiened by the one which is unit norm & achieves $||A||_2 = ||Av||_2$

      Note 1: The fact that $||A||_2$ is attenined at some v, $||v||_2 = 1$ is an instance of the extreme-value theorem for continuous function: $\mathbb{R}^n \to \mathbb{R}^m$ on closed & bounded sets.

      Note 2: $A : \mathbb{R}^n \to \mathbb{R}^m$ is continuous because $+, \cdot$ are themselves continuous. As well $||\ ||_2 : \mathbb{R}^m \to \mathbb{R}^+$ is also continuous because it's a square root of a sum of squares. So the composition $||A \cdot || : \mathbb{R}^n \to \mathbb{R}$ is a continuous function.

Proof

1. Let $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ be arbitrary with $||x||_2 = ||y||_2 = 1$, Then:

$$\begin{aligned} y^TAx &=< y, Ax > \\ &\leq ||y||_2||Ax||_2 \text{ by Cauchy-Schwarz} \\ &= ||Ax||_2 \text{ since } ||y||_2 = 1 \\ &\leq ||A|| \text{ by def of induced norm since } ||x||_2 = 1 \end{aligned}$$

   This shows $||A||$ is an upper bound for the set $y^TAx, ||x||_2 = ||y||_2 = 1$, since $||A||$, the induced norm, is bounded by: $\max_{||x||=1, x\in\mathbb{R}^n} ||Ax||_2$, which is larger than a regular $||Ax||_2$

   We now explicitly show that this upper bound is achieved. Take $v \in \mathbb{R}^n, ||v|| = 1, s.t. ||Av||_2 = ||A||$

   Take $u = \frac{Av}{||Av||_2}$, which is also unit, $||u||_2 = 1$

   Then $u^TAv = (\frac{Av}{||Av||_2})^TAv = \frac{||Av||_2^2}{||Av||_2} = ||Av||_2 = ||A||$

   with this choice of u, v i.e) $y = u, x = v$, we achieve the upper bound $||A||_2$

   This completes proof of first equal sign
2. $y^TAx = x^TA^Ty$ because both sides are scalar

3. This is a restatement of proof of 1 for A^T and an analogous proof applies.

Lemma 2

back to proof

With A, u, v as in **lemma 1** (u, v in the proof), let $\sigma_1 = ||A||_2$, then we have:

1. $Av = \sigma_1 u$
2. $A^T u = \sigma_1 v$

Proof

1. By construction of u, v from **lemma 1**:

$$u = \frac{Av}{||Av||_2}, ||Av||_2 = ||A||_2 = \sigma_1 \text{ by definition of both}$$
$$\text{then, we have: } \sigma_1 u = Av$$

2.
$$\begin{aligned}
\sigma_1 &= \sigma_1 u^T u, ||u||_2^2 = 1 \\
&= \sigma_1 u^T \frac{Av}{||Av||_2} \\
&= u^T Av \text{ since } \sigma_1 = ||A||_2 = ||Av||_2 \text{ (could've used lemma 1)} \\
&= v^T A^T u \text{ since transposing a scalar preserves the value} \\
&=< v, A^T u > \\
&\leq ||v||_2 ||A^T u||_2 \text{ by C-S} \\
&= ||A^T u||_2 \text{ because } ||v||_2 = 1 \\
&\leq ||A^T||_2 \text{ since } ||u||_2 = 1 \text{ by def of matrix 2-norm} \\
&= \sigma_1
\end{aligned}$$

so all inequalities are actually equalities (since σ_1 appears at the beginning and the end). Here in the C-S inequality, we know that equality occurs when the underlying vectors are collinear. Here $v \& A^T u$ are collinear (which was the point: i.e. to show that A^T maps u Back into $span\{v\}$). To figure out the multiplier, let $A^T u = \gamma v$ for some scalar γ

Then: $\sigma_1 = v^T A^T u = v^T \gamma v = \gamma$ since $||v^T v|| = ||v||_2^2 = 1$

This shows that u, v as defined are paris s.t.

- A maps v to $\sigma_1 u$
- A^T maps u to $\sigma_1 v$

These σ_1, u, v become the largest singular value & left & right singular vectors. The rest of the proof is by induction on the dimension of A where we remove $u_1 v$ from $\mathbb{R}^m \& \mathbb{R}^n$ respectively & reduce the subspace by looking at $u^\perp \& v^\perp$ and reduce the operator to a smaller one action on these subspaces. We now formalize

Main Proof of Existance of SVD

Let $A \in \mathbb{R}^{m \times n}$, assume $m \geq n$ because the case $m < n$ can be covered by transposition, since transposing an SVD form $u \Sigma v^T$, (u, v are orthogonal, Σ is diagonal, non negative) also produces an SVD form.

For one induction

Let

$$A = a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}_{m \times 1}$$

Then the set of vectors in $\mathbb{R}^1$ with $||v|| = 1$ is just $v = \pm 1$, with $||Av||_2 = ||a||_2$ or $||-a|| = ||a||_2$ for v s.t. $||v||_2 = 1$

Therefore, here $||A||_2 = ||a||_2 \overset{def}{=} \sigma_1$. Note: $||A||_2$ is the operator norm = vector 2-norm just for these $R^{m \times 1}$matrices

we have 2 choices of v_1 both achieving $||A||_2 = ||Av||$ we take $v = +1$ for convenience.

Then:

$$u = \frac{Av}{||Av||} = \frac{a}{||a||} = \frac{a}{\sigma_1}$$

Complete u to an orthonormal basis for $\mathbb{R}^m$ (we do not need to do for v because $v \in \mathbb{R}^1$, so it's already a complete basis)

Say $U = [u, u_2, \ldots, u_m]$ be an orthonormal basis for $\mathbb{R}^m$ (after completion). so U is an orthogonal matrix.

Then:

$$U_{m\times m}\begin{bmatrix}\sigma_1\\0\\\vdots\\0\end{bmatrix}_{m\times 1} v^T_{1\times 1} = [u, u_2, \ldots, u_m]\begin{bmatrix}\sigma_1\\0\\\vdots\\0\end{bmatrix} = \sigma_1 u = a = A$$

This is a valid SVD because

- v is orthogonal matrix in $\mathbb{R}^{1\times 1}$
- u is orthogonal matrix in $\mathbb{R}^{1\times 1}$ by construction in $\mathbb{R}^{m\times m}$, $\Sigma = \begin{bmatrix}\sigma_1\\0\\\vdots\\0\end{bmatrix}$ is diagonal ($m \times 1$ shape)

Induction Hypothesis

Suppose SVD can be constructed for any real matrix up to $dim(m-1)\times(n-1), n \geq 2$

Induction Step

Let $A \in \mathbb{R}^{m\times n}, m \geq n$. Take the following:

- $\sigma_1 = ||A||_2$ Note that σ_1 is already maximized based on lemma 1
- $v_1 \in \mathbb{R}^n$ s.t. $\sigma_1 = ||A||_2 = ||Av||_2$ with $||v_1||_2 = 1$
- $u_1 \in \mathbb{R}^m, u_1 = \frac{Av_1}{||Av_1||_2}$ so that $Av_1 = \sigma_1 u_1, A^T u_1 = \sigma_1 v_1$ (from lemma 2)

Complete both $v_1 \& u_1$ to orthonormal bases and write:

- $\tilde{V} = [v_2, \ldots v_n]$, and $[v_1, \tilde{V}]$ is orthogonal in $\mathbb{R}^n$
- $\tilde{U} = [u_2, \ldots u_m]$, and $[u_1, \tilde{U}]$ is orthogonal in $\mathbb{R}^m$

Apply $[u_1, \tilde{U}]^T$ and $[v_1, \tilde{V}]$ to the left and right of A respectively to get:

$$\begin{aligned}&[u_1, \tilde{U}]^T A [v_1, \tilde{V}]\\ &= \begin{bmatrix}u_1^T\\ \tilde{U}^T\end{bmatrix}[Av_1 \quad A\tilde{V}]\\ &= \left[\begin{array}{c|c} u_1^T A v_1 & u_1^T A\tilde{V} \\ \hline \tilde{U}^T A v_1 & \tilde{U}^T A \tilde{V}\end{array}\right]\end{aligned}$$

Shapes and Simplify:

- u_1^T : 1 x m, Av_1 : m x 1, thus the left top corner gives a scalar: 1 x 1
  - $u_1^T Av_1 = u_1^T \sigma_1 u_1 = \sigma_1$ by definition of u_1 and since $||u_1|| = 1$
- u_1^T : 1 x m, $A\tilde{V}$: m x (n-1), the shape matches, however, we can simplify it by:
  - $u_1^T A\tilde{V} = \sigma_1 v_1^T \tilde{V} = 0$ since $u_1^T A = (A^T u_1)^T = (\sigma_1 v_1)^T = \sigma_1 v_1^T$ and since $v_1 \perp \tilde{V}$, it ends up being 0
- $\tilde{U}^T$: (m-1) x m, Av_1 : m x 1, thus the left bottom corner matrix multiplication works
  - $\tilde{U} Av_1 = \tilde{U}\sigma_1 u_1 = 0$ since $Av_1 = u_1$ and $\tilde{U} \perp u_1$, thus it ends up being 0 (orthogonality)
- $\tilde{U}^T$: (m-1) x m, A : m x n, $\tilde{V}$: n x (n-1), thus the bottom right corner gives: (m-1) x (n-1)
  - $\tilde{U}^T A\tilde{V} \overset{def}{=} B \in \mathbb{R}^{(m-1)\times(n-1)}$. So by the induction hypothesis $\exists U_2, \Sigma_2, V_2$ in SVD form s.t. $B = U_2 \Sigma_2 V_2^T$

Rewrite:

$$[u_1, \tilde{U}]^T A [v_1, \tilde{V}] = \left[\begin{array}{c|c} u_1^T Av_1 & u_1^T A\tilde{V} \\ \hline \tilde{U}^T Av_1 & \tilde{U}^T A\tilde{V} \end{array}\right] = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & U_2\Sigma_2 V_2^T \end{array}\right]$$

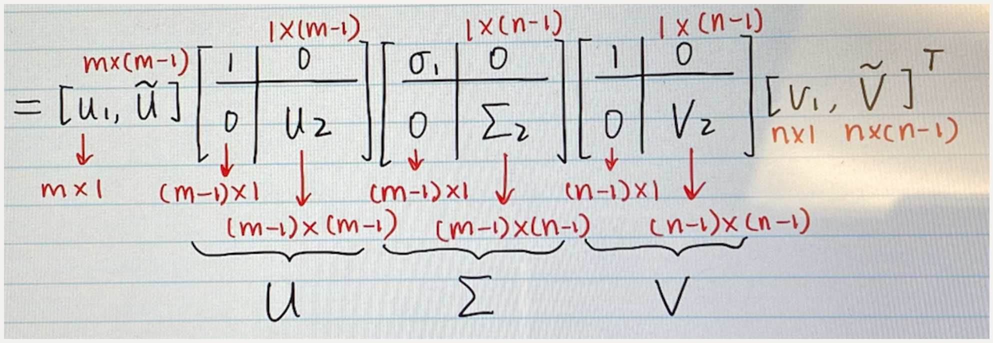
Recap from what has been covered so far in the proof:

- During the induction step, we reached a stage of showing that:
  $[u_1, \tilde{U}]^T A [v_1, \tilde{V}] = \left[\begin{array}{c|c} u_1^T Av_1 & u_1^T A\tilde{V} \\ \hline \tilde{U}^T Av_1 & \tilde{U}^T A\tilde{V} \end{array}\right] = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & U_2\Sigma_2 V_2^T \end{array}\right]$
  Where we invoked $I.H$ in B stating that it can be written in SVD form.
- Therefore:

  $$A = [u_1, \tilde{U}] \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & U_2\Sigma_2 V_2^T \end{array}\right] [v_1, \tilde{V}]^T$$

  we need to extract U_2, V_2^T from the middle matrix in order to obtain a diagonal

  Trick:



Here, the product composing U is of two orthogonal matrices and the product of orthogonal matrices is also an orthogonal matrix. (e.g. if Q_1, Q_2 are square, orthogonal of the same dimension, then $(Q_1Q_2)^T(Q_1Q_2) = Q_2^TQ_1^TQ_1Q_2 = I$)

Then we can write:

$$A = \left(\begin{bmatrix} u_1, \tilde{U} \end{bmatrix} \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & U_2 \end{array} \right] \right) \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array} \right] \left(\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array} \right] [v_1, \tilde{V}]^T \right) = U\Sigma V^T$$

shapes: (m x m * m x m) * m x n * (n x n * n x n) -> m x n

- **This completes the proof of existence!!**

To achieve ordering of $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_{min(m,n)}$, you can argue that the form achieved is already with the correct ordering since we can show that $\sigma_1 = ||A||_2$ by choice is the largest singular value & the others are correctly ordered by the induction hypothesis.

OR, we can argue that a correct reordering can be achieved by reordering the singular vectors.

Preserves:

1. orthogonality of left & right matrices
2. preserve the product: this can be seen most easily by recalling that SVD can be expanded (equivalent to full SVD) as a sum of rank-1 matrices.

$$\sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T$$

Helpful Notes for Assignment 4

1. Q3, d -> removed! Incorrect question

Lower Rank approximations to A

Def: Distance between matrices

Given a matrix norm $||\ \ ||$, the distance between matrices $A, B \in \mathbb{R}^{m \times n}$ is defined by $||A - B||$

A question you may ask is: what is the closest matrix B of rank k to A, where $rank(A) = n$ or $> k$

what is the answer in the 2-norm?

Answer (A4)

Take $A = \sum_{i=1}^{min(m,n)} \sigma_i u_i v_i^T$ (SVD), Take $B = \sum_{i=1}^{k} \sigma_i u_i v_i^T$. This works in both 2-norm and F-norms.

Another Note for A4

Given $A \in \mathbb{R}^{m \times n}, Q \in \mathbb{R}^{m \times m}$ orthogonal, then $||QA||_2 = ||A||_2$ because the operator norm is defined via a vector norm & the matrix Q preserves the underlying vector norm. Formally, we write:

$$||A||_2 = \max_{x \in \mathbb{R}^m} ||Ax||_2 = ||Av||_2 \text{ for some } v, ||v||_2 = 1$$

$$= ||QAv||_2 \text{ since Q is orthogonal matrix (recall: } ||Qu|| = ||u||)$$

since $||QAx||_2 = ||Ax||_2, \forall x, ||x||_2 = 1$, then $||QA||_2 = ||A||_2$

What about $||AQ_2||_2$ where $Q_2 \in \mathbb{R}^{n \times n}$?

- It is also orthogonal. Again, this is equal to $||A||_2$ since for any matrix, the 2-norm equal the 2-norm of its transpose

What about $||QA||_F$ & $||AQ_2||_F$?

Do these orthogonal matrices preserve the $||A||_F$?

$||A||_F^2 = ||a_1||_2^2 + \cdots + ||a_n||_2^2$ where $A = [a_1, \ldots a_n]$

Then we have:

$$\begin{aligned} ||QA||_F^2 &= ||[Qa_1, \ldots Qa_n]||_F^2 \\ &= ||Qa_1||_2^2 + \cdots + ||Qa_n||_2^2 \\ &= ||a_1||_2^2 + \cdots + ||a_n||_2^2 \text{ Q preserves 2-norm of vectors} \\ &= ||A||_F^2 \end{aligned}$$

So we have: $||A||_F = ||Q||_F$, as well as $||AQ_2||_F = ||A||_F$, since transposing a matrix preserves its F-norm (since we are just summing all elements squared & taking square root)

Topic 4: Eigenvalue Decomposition (EVD)

When discussing EVD, even for real matrices, we can not avoid $\mathbb{C}$ because generally characteristic polynomials only splitt to linear factors in $\mathbb{C}$.

In this class, we use $\mathbb{C}$ for some of the theory but only explore algorithms on subclasses of matrices (real, symmetric)

which are guaranteed to have real eigenvalues.

Definition: Eigen Pair

Let $A \in \mathbb{C}^{n\times n}$(or $\mathbb{R}^{n\times n}$). An eigenvalue of A is a scalar $\lambda \in \mathbb{C}$ s.t. $\exists v \in \mathbb{C}^n$ called eigenvector ($v \neq 0$), satisfying that $Av = \lambda v$. The pair (λ, v) is called an eigenpair of A.