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

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

       $V \rightarrow$ no need to break down.

       Then:

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

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

       $U \rightarrow$ leave unchanged

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

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

       We then get:

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

c. **<u>Case 3:</u>** SVD as a sum of rank-1 matrices

$U = [u_1, u_2. \dots u_m], u_i \in \mathbb{R}^m$

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

$V = [v_1, v_2, \dots v_n], v_j \in \mathbb{R}^n$

Then we have:

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

Definitions of SVD components

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

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

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. $\sigma_1 \geq \sigma_2 \geq \ldots . \geq \sigma_r > 0$

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

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

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

- 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
$Image(A) = span\{u_1, u_2, \ldots u_r\}$

Rank:

$col\ rank(A) = dim\ Image(A) = r = \#$ 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 $A^T = V\Sigma^T U^T$

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:

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

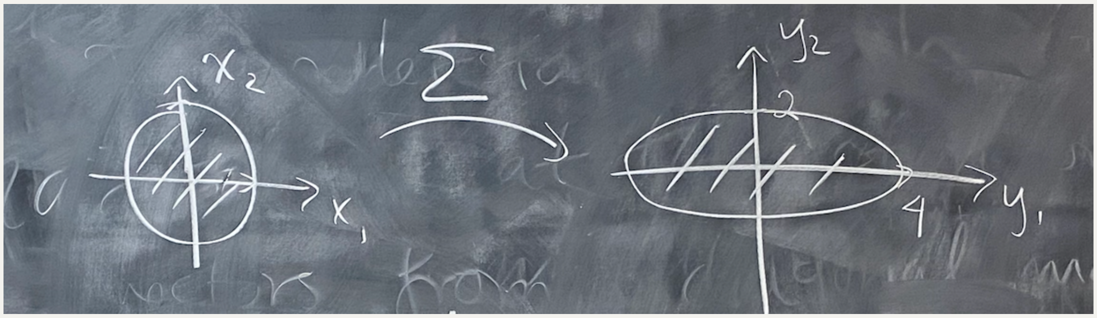
Then we have:

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

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 $1 \geq ||x||^2 = \ldots$

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:

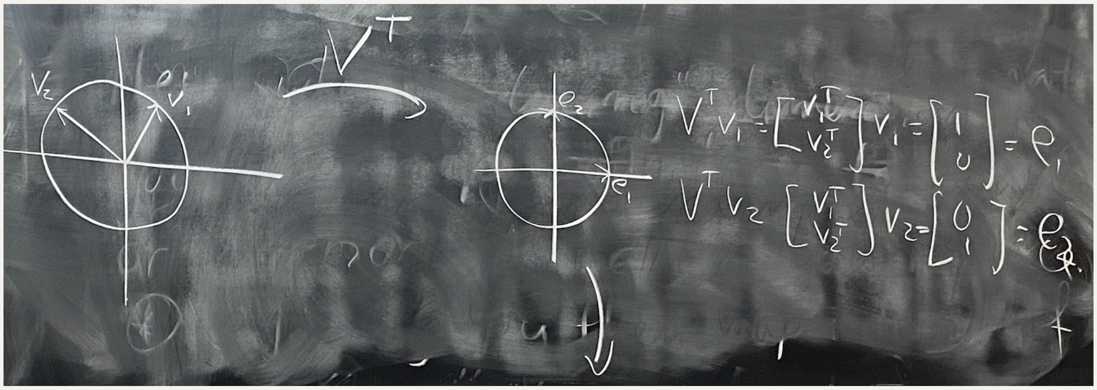
$$\begin{aligned} 1 &= x_1^2 + x_2^2 + x_3^2 \\ &\geq x_1^2 + x_2^2 \\ &= (\frac{y_1}{3})^2 + (\frac{y_2}{2})^2 \end{aligned}$$

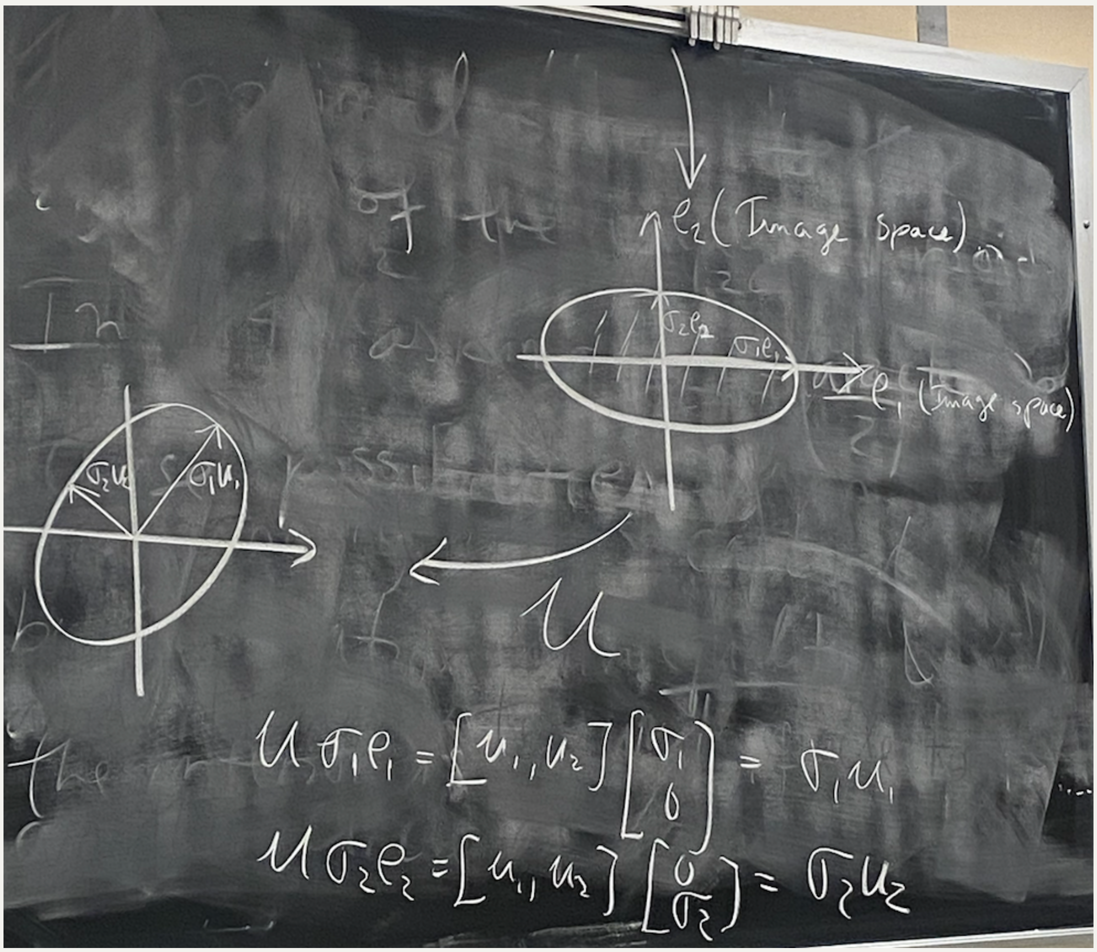
When “reducing” dimension, you may map boundary to boundary or interior. As well if you replace the equal sign by:

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

Now adding the other components of the SVD 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**

- **Def:** Given $A \in \mathbb{R}^{m \times n}$, the Frobenius norm is the following:

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

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

3. ***Side Question:***
   - Is the Frobenius norm induced? How do we know?

     The F-norm is not induce because:

     e.g)

     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

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

Then we have:

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

- where both $||A||_2, ||A^T||_2$ are induced

Proof

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

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

   This shows $||A||$ is an upper bound for the set $y^T Ax, ||x||_2 = ||y||_2 = 1$

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

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

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

   withi 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. $y^T Ax = x^T A^T y$ because both sides are scalar
3. This is a restatement of proof of 1 for A^T and an analogous proof applies.