Proven of Tensors: Revisiting the SVD

If \( A = U \Sigma V^T \) then \( \text{rank}(A) = r = \# \text{ of nonzero singular values } \sigma_1, \ldots, \sigma_r \)

**PF:** **Lemma:** If \( B \) is invertible then \( \text{rank}(AB) = \text{rank}(A) = \text{rank}(BA) \)

**PF:** **Remark:**
\[ \text{rank}(AB) \leq \text{rank}(A) \] by **LHW**

but:
\[ \text{rank}(A) = \text{rank}((AB)B^{-1}) \leq \text{rank}(AB) \]

\( \Rightarrow \text{rank}(A) = \text{rank}(AB) \) \( \checkmark \)

\( \Rightarrow \) likewise for \( \text{rank}(BA) \).

So:
\[ \text{rank}(A) = \text{rank}(U \Sigma V^T) \]
\[ = \text{rank}(\Sigma) \text{ since } U, V \text{ invertible} \]
\[ = r \text{ (why?)} \]
Recall: can write
\[ A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \]

This shows: \( \text{rank}(A) = \min \# \text{ of rank 1 matrices it takes to sum to } A \)

Why: so ofc, can't write \( A \) as less.

Then \( r = \text{rank}(A) \) rank 1 matrices

Since by HFD: \( \text{rank} \left( A_1 + \ldots + A_k \right) \)
\( \leq \text{rank}(A_1) + \ldots + \text{rank}(A_k) \)
\( \leq k \) if \( \# \) of all \( \text{rank}(A_i) = 1 \).

As above shows: we can achieve \( A \) as sum of \( r \) rank 1 matrices

This observation will allow us to generalize concept of rank to tensors.

As shows: if we defined \( \text{rank}(A) \) as least \# of rank 1 matrices needed to sum to \( A \) — then they would be equivalent to orig. defn (\# of indep. columns)
If we take initial segments:

\[ A_k = \sum_{\ell=1}^{k} u_\ell v_\ell^T \quad \text{(for } k \leq r) \]

Then:
1. \( A_k \) is rank \( k \) (why? — we knew SVD of \( A_k \) — why?)
2. \( A_k \) is "best rank \( k \) approximation to \( A \)" in a sense we will now make precise.

Define: \( L^2 \) norm of \( A \) is defined by:

\[
\| A \|_2 = \min_{\| X \|_2 = 1} \| AX \|_2 = \min_{\text{s.t. } \| X \|_2 = 1} \| AX \|_2
\]

= \( \sigma_1 \), first singular value of \( A \).

Sometimes write \( \| A \| \) for \( \| A \|_2 \).

We can show: \( \| \cdot \|_2 \) is norm on the space \( \mathbb{R}^{m \times n} \):

\[
\| A \|_2 \geq 0, \quad \| cA \|_2 = |c| \| A \|_2, \quad \| A + B \|_2 \leq \| A \|_2 + \| B \|_2, \quad \| AB \|_2 \leq \| A \|_1 \| B \|_\infty
\]
Theorem (Eckart-Young): If $B$ is rank $k$ ($k \leq r = \text{rank}(A)$) then:

$$\|A-B\|_2 \geq \|A-A_k\|_2 = \sigma_{k+1}$$

i.e. $A$ is closest (in $L^2$ norm) to $A_k$ among all rank $k$ matrices.

Proof: SPS not, i.e. SPS there is $B$ with $\text{rank}(B) = k$ such that $\|A-B\|_2 < \|A-A_k\|_2 = \sigma_{k+1}$

- By definition of $\|\cdot\|_2$, for any $\omega$ we have

$$\|A-B\|_2 \leq \sigma_{k+1} \Rightarrow \|A-B\|_2 \leq \sigma_{k+1} \|\omega\|_2$$

- By rank-nullity: nullspace $N(B)$ of $B$

  $u$ rank $n-k$

- For any $\omega \in N(B)$ we have $(A-B)\omega = Au$, so:

  $$\|A-B\|_2 \leq \|A\|_2 \|\omega\|_2 < \sigma_{k+1} \|\omega\|_2$$
- So \( NCB \) is an \((n-k)\)-dimensional subspace of \( \mathbb{R}^n \) s.t. \( \text{tri } NCB \)

\[ \|A\tilde{w}\| < \sigma_{k+1} \|\tilde{w}\| \]

- Now consider the subspace \( V \) spanned by \( \tilde{v}_1, \ldots, \tilde{v}_{k+1} \) -> the first \( k+1 \) singular vectors of \( A \).

- Thus space is \((k+1)\)-dim'ed and for every \( \tilde{w} \in V \) we have

\[ \|A\tilde{w}\| \geq \sigma_{k+1} \|\tilde{w}\| \]

Why? because we can write every such \( \tilde{w} \) as a linear combo of \( \tilde{v}_1, \ldots, \tilde{v}_{k+1} \) and these vectors are all stretched by a factor of \( \sigma_{k+1} \) (since \( \sigma_1 > \sigma_2 > \cdots > \sigma_{k+1} \)).

- Now: since \( \dim(NCB) + \dim(V) > n \) there must be some vectors in their intersection.
- But then
  \[ \| A w \|_2 < \| \tilde{w} \|_2 \quad \text{and} \quad \| A u \|_2 \geq \sigma_k, \| \tilde{u} \|_2 \quad \text{since} \quad \tilde{w} \in \mathcal{N}(B) \]
  \[ \text{and} \quad \| A u \|_2 \geq \sigma_{k+1}, \| \tilde{u} \|_2 \quad \text{since} \quad \tilde{u} \in \mathcal{V} \]

- Consequently

Summary SVD gives a way of
writing \( A \) as sum of \( r \) \((= \text{rank}(A))\)
many rank 1 matrices:

\[ A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \]

Given \( k \leq r \) the implied sum \( A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \) rank matrix that best
approximates \( A \) \((\text{wrt } L^2 \text{ norm})\) among
all rank \( k \) matrices.

- What about \( \text{wrt other norms?} \)
Define the Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$ defined by:

$$
\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}
$$

E.g. $\|\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}\|_F = \sqrt{1^2 + 2^2 + (-1)^2 + 0^2} = \sqrt{6}$

The Frobenius norm treats $A$ "like a linear combination of vectors".

One can check $\|\cdot\|_F$ also satisfies norm properties:

$$
\|A\|_F \geq 0, \quad \|cA\|_F = |c| \|A\|_F,
$$

$$
\|A + B\|_F \leq \|A\|_F + \|B\|_F
$$

Notice: If $Q$ is orthogonal, then since $\|Qx\|_2 = \|x\|_2$ for all vectors, and $\|A\|_F$ sums the norms of columns of $A$, $\|QA\|_F$ sums the norms of columns of $QA$ and $QA$ multiplies each column by $Q$ we must have $\|QA\|_F = \|A\|_F$.
They also gave us a way to write $\|A\|_F$ in terms of $A$'s singular values: if $A = U \Sigma V^T$

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

$$\text{or}
\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}
$$

Amazingly, $A_k = \sum_{l=1}^{k} \sigma_l u_l v_l^T$ is also the nearest rank $k$ matrix to $A$ in the Frobenius norm.

**Theorem (Eckart-Young for $\|\cdot\|_F$)**

If $B$ is rank $k$, then

$$\|A-B\|_F \geq \|A-A_k\|_F$$

$$= \sqrt{\sum_{k+1}^{\min\{m,n\}} \sigma_k^2}$$

**PF:** omitted.