Let A be an $n \times m$ matrix. We show the existence of the (full) SVD, i.e. there is an $n \times n$ orthogonal matrix U, an $n \times m$ diagonal matrix Σ with weakly decreasing entries, and an $m \times m$ orthogonal matrix V, such that $A = U \Sigma V^\top$.

First, we reduce our problem to finding an orthonormal basis of $\mathbb{R}^m$ whose property of orthogonality is preserved under the transformation defined by A.

Lemma 1. *Assume there exists an orthonormal basis $v_1, \ldots, v_m \in \mathbb{R}^m$ such that the vectors Av_i are pairwise orthogonal, meaning that $(Av_i) \cdot (Av_j) = 0$ for all $i \neq j$. Then A has an SVD.*

Proof. Some of the Av_i's may be zero. Reorder the v_i's so that those ones come last. Hence, for some r, we may assume that

- $Av_1, \ldots, Av_r$ are nonzero and of weakly decreasing length.
- $Av_{r+1}, \ldots, Av_m$ are zero.

Let V be the $m \times m$ orthogonal matrix built from $v_1, \ldots, v_m$. Let Σ be the $n \times m$ diagonal matrix whose first r diagonal entries are $\|Av_1\|, \ldots, \|Av_r\|$, and whose remaining entries are zero. Let U be an $n \times n$ orthogonal matrix whose first r columns are given by the orthonormal collection $\frac{1}{\|Av_1\|} Av_1, \ldots, \frac{1}{\|Av_r\|} Av_r$, and whose remaining columns are arbitrary. Then we have

$$Av_i = \|Av_i\| \, (i\text{-th column of } U)$$

for all $i = 1, \ldots, m$. Therefore $AV = U\Sigma$, so $A = U \Sigma V^\top$, as desired. □

Next, we show that this 'good' orthonormal basis exists.

Lemma 2. *Let $v \in \mathbb{R}^m$ be a unit vector which maximizes $\|Av\|^2$. Then, for any $w \in \mathbb{R}^m$ such that $w \cdot v = 0$, we have $(Aw) \cdot (Av) = 0$.*

Proof. The maximality property of v implies that $t = 0$ is a global maximum of the function

$$f(t) = \left\| A \left(\tfrac{v + tw}{\|v + tw\|} \right) \right\|^2,$$

because $\frac{v+tw}{\|v+tw\|}$ is a unit vector. Therefore, $f'(0) = 0$.

This derivative is computed as follows. First, note that

$$\begin{aligned} f(t) &= \tfrac{1}{\|v+tw\|^2} \, (v + tw)^\top A^\top A (v + tw) \\ &= \tfrac{1}{1 + t^2 \|w\|^2} \left(\|Av\|^2 + 2t \, (Aw) \cdot (Av) + t^2 \, \|Aw|^2 \right) \end{aligned}$$

where we have used that $\|v\| = 1$ and $v \cdot w = 0$. By looking at the t coefficient and ignoring higher powers of t, we see that $f'(0) = 2 \, (Aw) \cdot (Av)$. Therefore $(Aw) \cdot (Av) = 0$, as desired. □

Lemma 3. *Let A be an $n \times m$ matrix. There exists an orthonormal basis $v_1, \ldots, v_m \in \mathbb{R}^m$ such that the vectors Av_i are pairwise orthogonal.*

Proof. Proceed by induction on m. For the base case $m = 1$, just take v_1 to be any unit vector.

Assume $m \geq 2$. Let v_m be a unit vector[1] which maximizes $\|Av_m\|^2$. Consider the $(m-1)$-dimensional subspace $W \subset \mathbb{R}^m$ consisting of all vectors w such that $w \cdot v = 0$. By choosing an orthonormal basis of W, we obtain an $m \times (m-1)$ orthogonal matrix Q whose column space is W. Applying the inductive hypothesis to the $n \times (m-1)$ matrix AQ, we obtain orthonormal vectors $x_1, \ldots, x_{m-1} \in \mathbb{R}^{m-1}$ such that the AQx_i are pairwise orthogonal. Since Q is an orthogonal matrix, the vectors Qx_i are orthonormal.

Set $v_i = Qx_i$ for $i = 1, \ldots, m-1$. Then the $v_1, \ldots, v_{m-1}$ are orthonormal, and $Av_1, \ldots, Av_{m-1}$ are pairwise orthogonal. Since $v_1, \ldots, v_{m-1} \in W$ by construction, the set $v_1, \ldots, v_m$ is an orthonormal basis, and Lemma 2 implies that Av_m is orthogonal to the $Av_1, \ldots, Av_{m-1}$. Therefore $v_1, \ldots, v_m$ has the desired properties, so the inductive step is proved. □

[1] Such a v_m exists because any continuous function on a compact domain attains its maximum. Indeed, a unit vector v varies on the unit sphere in $\mathbb{R}^m$, which is compact, and the map $v \mapsto \|Av\|^2$ is continuous.