# Lecture 4 - *The Singular Value Decomposition (SVD)*

**OBJECTIVE:**
The SVD is a matrix factorization that has many applications:
e.g., information retrieval, least-squares problems, image processing
It is so important that we introduce it at this early stage.

◇ GEOMETRIC INTERPRETATION

The SVD is motivated by the following geometric fact:

*The image of the unit sphere under any $m \times n$ matrix is a hyperellipse.*

**Note 1.** *An "ellipse" is a 2D object* $\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right)$.

*In 3D, it is called an "ellipsoid"* $\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right)$.

*The generalization to $m$ dimensions is called a hyper-ellipse.*

(Even the unit sphere in $m$ dimensions is more accurately called a hypersphere!)

The sphere maps into the hyperellipse by stretching the sphere by (possibly zero) factors

$$\sigma_1, \sigma_2, \ldots, \sigma_m$$

in orthogonal directions

$$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \qquad \in \mathbb{R}^m.$$

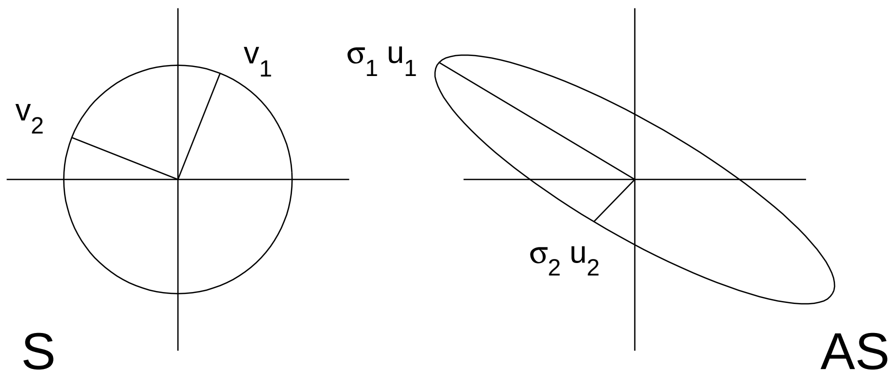
It is convenient to take the $\mathbf{u}_i$ to be unit vectors. Then, the vectors $\{\sigma_i \mathbf{u}_i\}_{i=1}^m$ can be viewed as the *principal semi-axes* of the hyperellipse with lengths

$$\sigma_1, \sigma_2, \ldots, \sigma_m$$

If $\mathbf{A}$ has rank $r$, then exactly $r$ of the lengths $\sigma_i$ will be nonzero. In particular, if $m \geq n$, then at most $n$ of the $\sigma_i$ will be nonzero.

These facts are not obvious, but for now just assume they are true!

Here is the basic picture:




Let $\mathbf{S}$ be the unit sphere in $\mathbb{R}^n$.

Choose any $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$.

For simplicity, let $\text{rank}(\mathbf{A}) = n$ (full rank).

The image $\mathbf{AS}$ is a hyperellipse in $\mathbb{R}^m$ with the following properties:

1. The $n$ *singular values* $\sigma_1, \sigma_2, \ldots, \sigma_n$ of $\mathbf{A}$ are the lengths of the principal semi-axes of $\mathbf{AS}$. By convention, we assume the $\sigma_i$ are ordered in descending order.
   i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$
   (why $\sigma_i > 0$?)

2. The set of unit vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ are the $n$ *left singular vectors* of $\mathbf{A}$. They are the directions of the principal semi-axes of $\mathbf{AS}$ corresponding to the $n$ singular values.
   $\therefore \sigma_i \mathbf{u}_i$ is the $i^{\text{th}}$ largest principal semi-axes of $\mathbf{AS}$.

3. The set of unit vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are the $n$ *right singular vectors* of $\mathbf{A}$. They are the "pre-images" of the $\mathbf{u}_i$ numbered accordingly.
   i.e., $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$
   (why "left" and "right" will be explained below)

## $\diamond$ REDUCED SVD

The equations relating the right singular values $\{\mathbf{v}_j\}$ and the left singular vectors $\{\mathbf{u}_j\}$ are

$$\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \qquad j = 1, 2, \ldots, n$$

i.e.,

$$[\mathbf{A}] \left[\ \mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n\ \right] = \left[\ \mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_n\ \right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

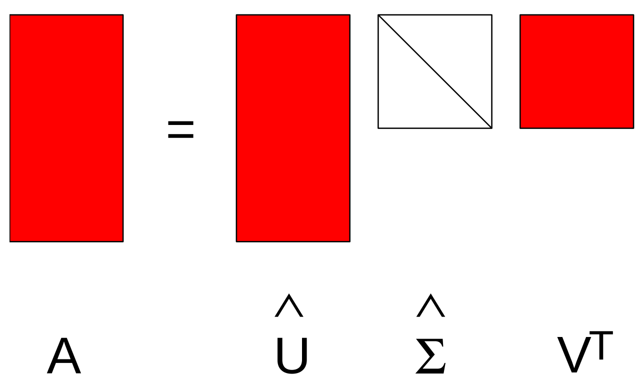
or

$$\mathbf{A}\mathbf{V} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}$$

**Note 2.** $\hat{\mathbf{\Sigma}} \in \mathbb{R}^{n \times n}$ *is diagonal with positive entries.*
$\hat{\mathbf{U}} \in \mathbb{R}^{m \times n}$ *with orthonormal columns.*
$\mathbf{V} \in \mathbb{R}^{n \times n}$ *with orthonormal columns.*
*(*$\Rightarrow \mathbf{V}$ *is orthogonal so* $\mathbf{V}^{-1} = \mathbf{V}^T$*)*

$$\therefore \qquad \mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\mathbf{V}^T$$

This factorization is called the *reduced SVD* of $\mathbf{A}$.
(We'll see why "reduced" in a minute.)



$\diamond$ FULL SVD

In practice, SVD is mostly used in its reduced form. But, theoretically, there is a "full SVD" that is more standard.

The idea of the full SVD is as follows:

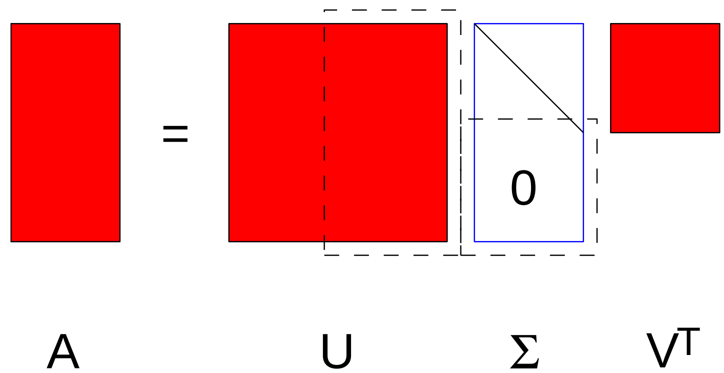
$\hat{\mathbf{U}}$ is made up of $n$ orthonormal vectors in $\mathbb{R}^m$.
So, unless $m = n$, this cannot be a basis for $\mathbb{R}^m$, and $\hat{\mathbf{U}}$ cannot be orthogonal.
But, we can make $\hat{\mathbf{U}}$ orthogonal by adding $m - n$ additional orthonormal columns.
Let's suppose we do this and call the result $\mathbf{U}$.

But, if we replace $\hat{\mathbf{U}}$ by $\mathbf{U}$, we will have to change $\hat{\mathbf{\Sigma}}$ as well
$\rightarrow$ for the factorization not to change, we must multiply the last $m - n$ columns of $\mathbf{U}$ by 0.

This leads to the full SVD

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{V} \in \mathbb{R}^{n\times n}$ are both orthogonal, and $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$.




In this framework, we no longer need to assume $\mathbf{A}$ has full rank.

If $\text{rank}(\mathbf{A}) = r < n$, only $r$ singular vectors of $\mathbf{A}$ are determined by the geometry of the hyperellipse and we will have to add $m - r$ orthonormal columns to $\mathbf{U}$ and $n - r$ orthonormal columns to $\mathbf{V}$. $\mathbf{\Sigma}$ will have $r$ positive diagonal entries.

Alternatively, in this case it is possible to further reduce the reduced SVD of $\mathbf{A}$ by letting $\hat{\mathbf{U}} \in \mathbb{R}^{m \times r}$, $\hat{\mathbf{\Sigma}} \in \mathbb{R}^{r \times r}$, and $\hat{\mathbf{V}}^T \in \mathbb{R}^{r \times n}$.

## $\diamond$ FORMAL DEFINITION OF SVD

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, the SVD of $\mathbf{A}$ is a factorization of the form

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbf{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal, and $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ is "diagonal".

**Note 3.** *1. There is no assumption that $m \geq n$ or that $\mathbf{A}$ has full rank.*

*2. All diagonal elements of $\mathbf{\Sigma}$ are non-negative and in non-increasing order:*

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$$

*where $p = \min(m, n)$*

*3. $\mathbf{\Sigma}$ has the same shape as $\mathbf{A}$, but $\mathbf{U}$, $\mathbf{V}$ are square.*

## $\diamond$ EXISTENCE AND UNIQUENESS OF THE SVD

**Theorem 1.** *Every matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ has singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.*
*Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined.*
*If $\mathbf{A}$ is square and $\sigma_i \neq \sigma_j$ for all $i \neq j$, the left singular vectors $\{\mathbf{u}_j\}$ and the right singular vectors $\{\mathbf{v}_j\}$ are uniquely determined to within a factor of $\pm 1$.*