1 Singular Value Decomposition & Basic Applications (4.1, 4.2)

Theorem 1.1: Geometric Singular Value Decomposition Theorem

Let $A \in \mathbb{R}^{n\times m}$ be a nonzero matrix with rank r. Then, $\mathbb{R}^m$ has an orthogonal basis $v_1, v_2, \ldots, v_m$, $\mathbb{R}^n$ has an orthonormal basis $u_1, u_2, \ldots, u_n$, and there exists $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ such that

$$Av_i = \begin{cases} \sigma_i u_i & i = 1, \ldots, r \\ \mathbf{0} & i = r+1, \ldots, m \end{cases} \qquad A^T u_i = \begin{cases} \sigma_i v_i & i = 1, \ldots, r \\ \mathbf{0} & i = r+1, \ldots, n \end{cases}.$$

For SVD, there exists orthogonal matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

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

with

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_r & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Remarks:

- This is called a full SVD[1] Here, $\sigma_1, \sigma_2, \ldots, \sigma_r$ are called the *singular values*.
- Note that we can write U and V as a vector of vectors,

$$U = \left[u_1, u_2, \ldots, u_n\right], u_i \in \mathbb{R}^n$$

and

$$V = \left[v_1, v_2, \ldots, v_m\right], v_m \in \mathbb{R}^m.$$

1.1 Intuition

To get some intuition, let's suppose we start with $A = U\Sigma V^T$. Then, we know that

$$AV = U\Sigma V^T V \implies AV = U\Sigma.$$

Rewriting V and U as columns, we have

$$A \begin{bmatrix} v_1 & v_2 & \ldots & v_m \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_r & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix}. \tag{1}$$

For some matrix

$$B = \begin{bmatrix} b_1 & b_2 & \ldots & b_m \end{bmatrix},$$

[1] Later, we will introduced a reduced SVD.

then

$$B\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}b_1 & b_2 & \dots & b_m\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}\sigma_1 b_1 & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix}.$$

Likewise,

$$B\begin{bmatrix}0 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}b_1 & b_2 & \dots & b_m\end{bmatrix}\begin{bmatrix}0 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \sigma_2 b_2 & \dots & \mathbf{0}\end{bmatrix}.$$

Notice how σ_i scales the ith column of B. Now, let's suppose we combine the operations above into one matrix:

$$B\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}0 & b_2 & \dots & 0\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}\sigma_1 b_1 & \sigma_2 b_2 & \dots & \mathbf{0}\end{bmatrix}.$$

Here, both column 1 and 2 are scaled. So, going back to equation (1), we have

$$\begin{bmatrix}\sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix}.$$

This gives us the formula,

$$Av_i = \begin{cases}\sigma_i u_i & i = 1,\dots,r\\ \mathbf{0} & i = r+1,\dots,m\end{cases}.$$

Likewise, let's consider A^T;

$$\begin{aligned}A^T &= (U\Sigma V^T)^T\\ &= (V^T)\Sigma^T U^T\\ &= V\Sigma U^T.\end{aligned}$$

From there, we have $A^T U = V\Sigma^T U^T U = V\Sigma^T$. Expanding out the matrices, we have

$$A^T\begin{bmatrix}u_1 & u_2 & \dots & u_n\end{bmatrix} = \begin{bmatrix}v_1 & v_2 & \dots & v_m\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & \sigma_r & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} = \begin{bmatrix}\sigma_1 v_1 & \sigma_2 v_2 & \dots & \sigma_r v_r & \mathbf{0} & \dots\end{bmatrix}.$$

Note that Σ^T is not a tall matrix, but a wide one; instead of being $n \times m$, it's $m \times n$. This gives us the formula

$$A^T u_i = \begin{cases}\sigma_i v_i & i = 1,\dots,r\\ \mathbf{0} & i = r+1,\dots,m\end{cases}.$$

In particular, we can see that

$$
\begin{aligned}
v_1 &\xrightarrow[\sigma_1]{A} u_1 & \qquad u_1 &\xrightarrow[\sigma_1]{A^T} v_1\\
v_2 &\xrightarrow[\sigma_2]{} u_2 & u_2 &\xrightarrow[\sigma_2]{} v_2\\
v_3 &\xrightarrow[\sigma_3]{} u_3 & u_3 &\xrightarrow[\sigma_3]{} v_3\\
&\vdots & &\vdots\\
v_r &\xrightarrow[\sigma_r]{} u_r & u_r &\xrightarrow[\sigma_r]{} v_r\\
v_{r+1} &\to \mathbf{0} & u_{r+1} &\to \mathbf{0}\\
&\vdots & &\vdots\\
v_m &\to \mathbf{0} & u_n &\to \mathbf{0}
\end{aligned}
$$

So, in particular, the range of matrix A, $\mathcal{R}(A)$, is spanned by

$$
\begin{aligned}
v_1 &\xrightarrow[\sigma_1]{A} u_1\\
v_2 &\xrightarrow[\sigma_2]{} u_2\\
v_3 &\xrightarrow[\sigma_3]{} u_3\\
&\vdots\\
v_r &\xrightarrow[\sigma_r]{} u_r.
\end{aligned}
$$

The null space of A, $\mathcal{N}(A)$, is spanned by

$$
\begin{aligned}
v_{r+1} &\to \mathbf{0}\\
&\vdots\\
v_m &\to \mathbf{0}.
\end{aligned}
$$

For A^T, this is analogous.

1.2 Fundamental Subspaces

The SVD displays orthogonal bases for the four Fundamental subspaces, $\mathcal{R}(A)$, $\mathcal{N}(A)$, $\mathcal{R}(A^T)$, and $\mathcal{N}(A^T)$, where $\mathcal{R}$ is the *range* and $\mathcal{N}$ is the null space.

$$\mathcal{R}(A) = \operatorname{span}\{u_1, u_2, \dots, u_r\}.$$

$$\mathcal{N}(A) = \operatorname{span}\{v_{r+1}, v_{r+2}, \dots, v_m\}.$$

$$\mathcal{R}(A^T) = \operatorname{span}\{v_1, v_2, \dots, v_r\}.$$

$$\mathcal{N}(A^T) = \operatorname{span}\{u_{r+1}, u_{r+2}, \dots, u_n\}.$$

In particular,

$$\mathcal{R}(A) + \mathcal{N}(A^T) = \operatorname{span}\{u_1, u_2, \dots, u_n\}.$$

$$\mathcal{R}(A^T) + \mathcal{N}(A) = \operatorname{span}\{v_1, v_2, \dots, v_m\}.$$

We can see that $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$ and $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$.

Corollary 1.1

Let $A \in \mathbb{R}^{n\times m}$. Then, $\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = m$ and $\dim(\mathcal{R}(A^T)) + \dim(\mathcal{N}(A^T)) = n$.

1.3 Reduced SVD

We'll introduce this section with an example.

(Example.) Suppose we have a 3×3 matrix of rank[a] 2. Then,

$$A = \underbrace{\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}}_{V^T}$$
$$= \begin{bmatrix} \sigma_1 u_{11} & \sigma_2 u_{12} & 0 \\ \sigma_1 u_{21} & \sigma_2 u_{22} & 0 \\ \sigma_1 u_{31} & \sigma_2 u_{32} & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}.$$

As a result, we'll end up multiplying v_{13}, v_{23}, v_{33} by 0. So, instead, what if we have:

$$A = \underbrace{\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}_{\hat{\Sigma}} \underbrace{\begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \end{bmatrix}}_{\hat{V^T}}.$$

This is known as the reduced SVD.

[a] It has 3 rows but only has 2 non-zero singular values, σ_1 and σ_2

Theorem 1.2: Condensed SVD Theorem

Let $A \in \mathbb{R}^{n \times m}$ be a nonzero matrix of rank r. Then, there exists $\hat{U} \in \mathbb{R}^{n \times r}$, $\hat{\Sigma} \in \mathbb{R}^{r \times r}$, and $\hat{V} \in \mathbb{R}^{m \times r}$, such that $\hat{U}$ and $\hat{V}$ are isometries, and $\hat{\Sigma}$ is a diagonal matrix with main-diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ and

$$A = \hat{U}\hat{\Sigma}\hat{V^T}.$$

1.3.1 Relationship to Norm and Condition Number

Recall that we defined the matrix 2-norm as

$$||A|| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||_2}{||\mathbf{x}||} = \sigma_1,$$

where σ_1 is the largest singular value. Note that this definition also makes sense for $A \in \mathbb{R}^{n \times m}$.

Theorem 1.3

$$||A||_2 = \sigma_1.$$

Proof. In next lecture. □

Since A and A^T have the same singular values, we have the following corollary.

Corollary 1.2

$$||A||_2 = ||A^T||_2.$$