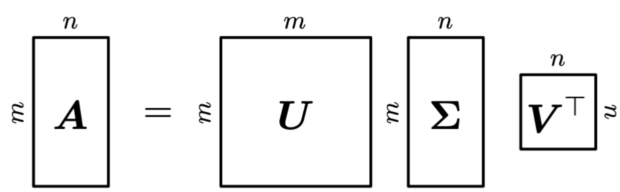
Contents

1 Singular Value Decomposition

> Every matrix can be written as a matrix product of a rotation matrix, a stretch, which is just a matrix with nonzero entries on some diagonal, and another rotation matrix.
>
> *Gilbert Strang*

Definition 1.1. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a rectangular matrix of rank $r \in [0, \min(m, n)]$. The SVD of $\boldsymbol{A}$ is a decomposition of the form:



- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix with column vectors $\boldsymbol{u}_i, i = 1, \ldots, m$
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix with column vectors $\boldsymbol{v}_j, j = 1, \ldots, n$
- $\boldsymbol{\Sigma}$ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geqslant 0$ and $\Sigma_{ij} = 0, i \neq j$
- The diagonal entries $\sigma_i, i = 1, \ldots, r$ (where $r \in [0, \min(m, n)]$), of $\boldsymbol{\Sigma}$ are called the singular values
- $\boldsymbol{u}_i$ are called the left-singular vectors and $\boldsymbol{v}_j$ are called the right-singular vectors.
- by convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_r \geqslant 0$

1.1 Singular Value Matrix Is Unique

The singular value matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is unique and rectangular. It is of the same size as $\boldsymbol{A}$. This means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding.

1. If $m > n$ (more rows than columns) for long matrix $\boldsymbol{A}$, then the matrix $\boldsymbol{\Sigma}$ has diagonal structure up to the n^{th} row and then consists of $0^\top$ row vectors from row $n + 1$ to row m below.

$$\boldsymbol{\Sigma} = \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix}}_{m \times n}$$

2. If $m < n$ (more columns than rows), the wide matrix $\mathbf{\Sigma}$ has a diagonal structure up to the m^{th} column and then consist of columns of 0's from column $m+1$ to column n :

$$\mathbf{\Sigma} = \underset{m \times n}{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}}$$

Remark. The SVD exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, whereas the eigen-decomposition applies to square matrices.

1.2 Alternative Representation

Another representation of the SVD is of the form:

$$\begin{aligned} A &= U\Sigma V^T \\ &= [\vec{u}_1 \cdots \vec{u}_r \cdots] \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ 0 & & & & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_r^T \\ \vdots \end{bmatrix} \\ &= [\vec{u}_1 \cdots \vec{u}_r \cdots] \begin{bmatrix} \sigma_1 \vec{v}_1^T \\ \vdots \\ \sigma_r \vec{v}_r^T \\ 0 \end{bmatrix} \\ &= \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T \end{aligned}$$

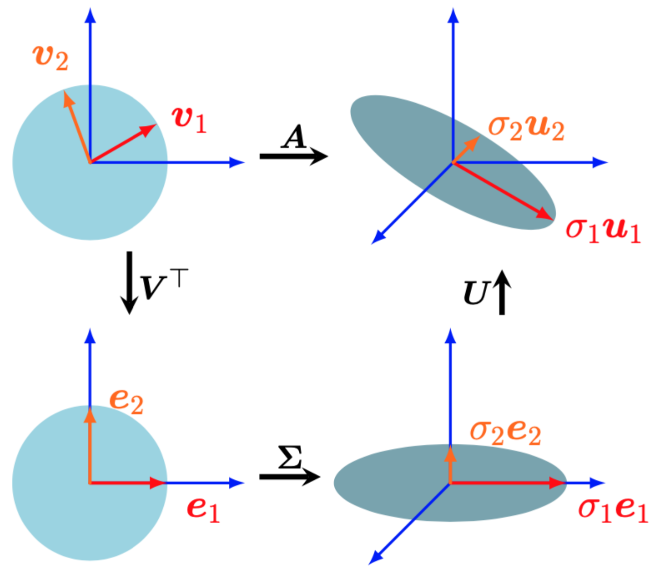
1.3 Geometric Intuition

The SVD of a matrix can be interpreted as a decomposition of a corresponding linear mapping $\Phi : \mathbb{R}^n(\text{Domain}) \to \mathbb{R}^m(\text{Co-domain})$ into three operations.

$$\underset{m \times n}{\boldsymbol{A}} = \underset{m \times m}{\boldsymbol{U}} \underset{m \times n}{\mathbf{\Sigma}} \underset{n \times n}{\boldsymbol{V}^\top}$$

The geometric intuition behind the SVD of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ boils down to sequential transformations. An example in $\boldsymbol{A} \in \mathbb{R}^{3 \times 2}$:

1. **Top-left to bottom-left**: $\boldsymbol{V}^\top$ performs a basis change in $\mathbb{R}^2$
2. **Bottom-left to bottom-right**: $\mathbf{\Sigma}$ scales and maps from $\mathbb{R}^2$ to $\mathbb{R}^3$; the ellipse in the bottom-right lives in $\mathbb{R}^3$ where the third dimension is orthogonal to the surface of the elliptical disk
3. **Bottom-right to top-right**: U performs a basis change within $\mathbb{R}^3$



1.4 Symbolic Intuition

Let an $m \times n$ matrix be the transformation matrix of a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases B and C of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Further, assume a second basis $\tilde{B}$ of $\mathbb{R}^n$ and $\tilde{C}$ of $\mathbb{R}^m$. Then

1. The matrix V performs a basis change in the domain $\mathbb{R}^n$ from $\tilde{B}$ (represented by the red and orange vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$) to the standard basis $\boldsymbol{B}$; the columns of matrix V are the $\tilde{B}$ basis vectors relative to the coordinates of the standard basis (top-left of the figure). $\boldsymbol{V}^\top = \boldsymbol{V}^{-1}$ performs a basis change from the standard basis B to $\tilde{B}$. The red and orange vectors are now aligned in the $\tilde{B}$ basis coordinate system in the bottom-left compared to the top-left of the figure, where they were represented with respect to the standard basis $\boldsymbol{B}$.

2. Having changed the coordinate system to new basis $\tilde{B}$, the transformation matrix $\boldsymbol{\Sigma}$ scales the new coordinates by the singular values σ_i (and adds or deletes dimensions), i.e., $\boldsymbol{\Sigma}$ is the transformation matrix of Φ with respect to bases $\tilde{B}$ and $\tilde{C}$, represented by the red and orange vectors being stretched and lying in the $\boldsymbol{e}_1 - \boldsymbol{e}_2$ plane (a space spanned by the $\tilde{B}$ basis vectors), which is now embedded in a third dimension in the bottom-right of the figure above (a space spanned by the $\tilde{C}$ basis vectors).

3. $\boldsymbol{U}$ performs a basis change in the codomain $\mathbb{R}^m$ from $\tilde{C}$ into the original basis of $\mathbb{R}^m$, represented by a rotation of the red and orange vectors out of the $e_1 - e_2$ plane. This is shown in the top-right of figure above.

Remark. The SVD expresses a change of basis in both the domain and codomain. This is in contrast with the **eigen-decomposition that operates within the same vector space**, where the same basis change is applied and then undone. What makes the SVD special is that these two different bases are simultaneously linked by the singular value matrix $\boldsymbol{\Sigma}$.

1.4.1 Example: $\boldsymbol{A} \in \mathbb{R}^{3\times 2}$

Consider a mapping of a square grid of vectors $\mathcal{X} \in \mathbb{R}^2$ that fit in a box of size 2×2 centered at the origin. Using the standard basis, we map these vectors using

$$
\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \\
&= \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}
\end{aligned}
$$

- We start with a set of vectors $\mathcal{X}$ (colored dots; see top-left panel of figure below) arranged in a grid
- We then apply $V^\top \in \mathbb{R}^{2\times 2}$, which rotates $\mathcal{X}$. The rotated vectors are shown in the bottom-left panel of figure
- Next, we map these vectors using the singular value matrix $\boldsymbol{\Sigma}$ to the codomain $\mathbb{R}^3$ (see the bottom-right panel). Note that all vectors lie in the $x_1 - x_2$ plane. The third coordinate is always 0. The vectors in the $x_1 - x_2$ plane have also been stretched by the singular values
- Lastly, $\boldsymbol{U}$ performs a rotation within the codomain $\mathbb{R}^3$ so that the mapped vectors are no longer restricted to the $x_1 - x_2$ plane; they still are on a plane as shown in the top-right panel of the figure. But they appear to be on an arbitrary plane in three-dimensional space.
- The direct mapping of the vectors $\mathcal{X}$ by $\boldsymbol{A}$ to the codomain $\mathbb{R}^3$ equals the sequence of transformation of $\mathcal{X}$ by $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$

1.5 Derivation of SVD

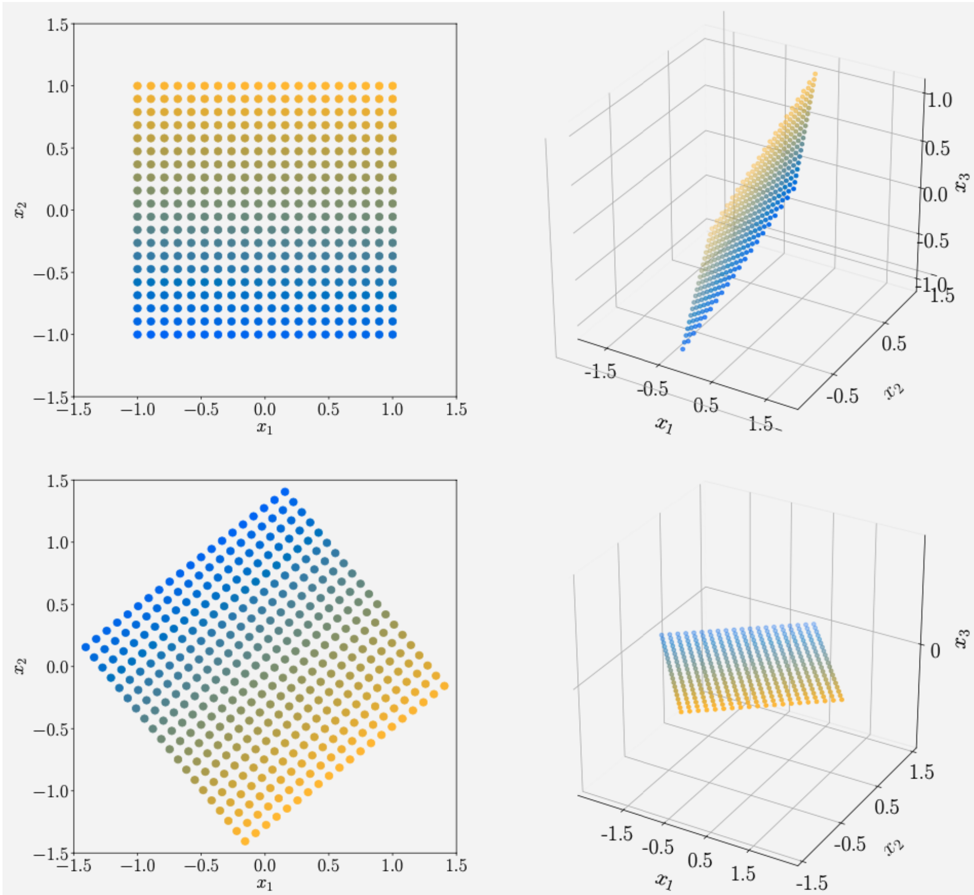
Computing the SVD of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is equivalent to finding two sets of orthonormal bases $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ (the left-singular vectors) and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ (the right-singular vectors) of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases, we will construct the matrices U and V.

1.5.1 Right Singular Vectors

1. Recall the following theorem regarding rectangular matrices.

   Definition 1.2. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we can always obtain a symmetric, positive semidefinite matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ by defining

   $$\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$$



Remark. If $\text{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, positive definite.

2. Using the theorem above, we can construct a square matrix from the product of $\boldsymbol{A}$ and its transpose, and find the eigen-decomposition of the resultant square matrix:

$$\underset{n\times m}{\boldsymbol{A}^\top}\underset{m\times n}{\boldsymbol{A}} = \boldsymbol{PDP}^\top = \underset{n\times n}{\boldsymbol{P}} \underset{n\times n}{\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}} \underset{n\times n}{\boldsymbol{P}^\top} \tag{1}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$. Let us assume the SVD of A exists. This means that

$$\underset{n\times m}{\boldsymbol{A}^\top}\underset{m\times n}{\boldsymbol{A}} = \left(\boldsymbol{U\Sigma V}^\top\right)^\top \left(\boldsymbol{U\Sigma V}^\top\right) = \boldsymbol{V\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U\Sigma V}^\top,$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\underset{n\times m}{\boldsymbol{A}^\top}\underset{m\times n}{\boldsymbol{A}} = \boldsymbol{V}\underset{n\times m}{\boldsymbol{\Sigma}^\top}\underset{m\times n}{\boldsymbol{\Sigma}}\boldsymbol{V}^\top = \underset{n\times n}{\boldsymbol{V}} \underset{n\times n}{\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}} \underset{n\times n}{\boldsymbol{V}^\top} \tag{2}$$

Comparing now equation 1 and equation 2, we see that the following are equivalent

$$\underset{n\times n}{\boldsymbol{V}} \underset{n\times n}{\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}} \underset{n\times n}{\boldsymbol{V}^\top} = \underset{n\times n}{\boldsymbol{P}} \underset{n\times n}{\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}} \underset{n\times n}{\boldsymbol{P}^\top}$$

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top,$$

$$\sigma_i^2 = \lambda_i$$

3. The eigenvectors of the square matrix $\boldsymbol{A}^\top\boldsymbol{A}$ that compose orthogonal matrix $\boldsymbol{P}$ are the right-singular vectors $\boldsymbol{V}$ of $\boldsymbol{A}$.

4. The eigenvalues of the square matrix $\boldsymbol{A}^\top\boldsymbol{A}$ that are the diagonal entries of matrix $\boldsymbol{D}$ are the squared singular values of $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$.

1.5.2 Left Singular Vectors

For the left-singular vectors U, we use the symmetric matrix $\underset{m\times n}{\boldsymbol{A}}\ \underset{n\times m}{\boldsymbol{A}^\top} \in \mathbb{R}^{m\times m}$ (instead of the previous $\boldsymbol{A}^\top\boldsymbol{A} \in \mathbb{R}^{n\times n}$).

1. The eigen-decomposition of the matrix $\underset{m\times n}{\boldsymbol{A}}\ \underset{n\times m}{\boldsymbol{A}^\top}$ is as follows:

$$\underset{m\times n}{\boldsymbol{A}}\ \underset{n\times m}{\boldsymbol{A}^\top} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \underset{m\times m}{\boldsymbol{P}} \underset{m\times m}{\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}} \underset{m\times m}{\boldsymbol{P}^\top}$$

2. Using the SVD of $\boldsymbol{A}$ and the fact that $\boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}$

$$\begin{aligned} \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times m}{\boldsymbol{A}^\top} &= \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top \\ &= \boldsymbol{U} \underset{m\times n}{\boldsymbol{\Sigma}}\ \underset{n\times m}{\boldsymbol{\Sigma}^\top} \boldsymbol{U}^\top \\ &= \underset{m\times m}{\boldsymbol{U}} \underset{m\times m}{\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix}} \underset{m\times m}{\boldsymbol{U}^\top} \end{aligned}$$

3. We see that

$$\underset{m\times m}{\boldsymbol{U}} \underset{m\times m}{\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix}} \underset{m\times m}{\boldsymbol{U}^\top} = \underset{m\times m}{\boldsymbol{P}} \underset{m\times m}{\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}} \underset{m\times m}{\boldsymbol{P}^\top}$$

4. The spectral theorem tells us that $\boldsymbol{A}A^{\top} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{\top}$ can be diagonalized and we can find an ONB of eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\top}$, which are collected in $\boldsymbol{S}$.

5. The eigenvectors of the square matrix $\underset{m\times n}{\boldsymbol{A}}\,\underset{n\times m}{\boldsymbol{A}^{\top}}$ that compose orthogonal matrix $\boldsymbol{P}$ (this time a $m \times m$ matrix instead of $n \times n$) are the left-singular vectors $\boldsymbol{U}$ of $\boldsymbol{A}$.

6. The eigenvalues of the square matrix $\underset{m\times n}{\boldsymbol{A}}\,\underset{n\times m}{\boldsymbol{A}^{\top}}$ are still the squared singular values of $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$. This time we have m of them in the co-domain rather than n of them in the domain.

1.5.3 Combining Right and Left Singular Vectors

We can find $\boldsymbol{V}$ and $\boldsymbol{U}$ by finding the eigen-decomposition of $\underset{n\times m}{\boldsymbol{A}^{\top}}\,\underset{m\times n}{\boldsymbol{A}}$ and $\underset{m\times n}{\boldsymbol{A}}\,\underset{n\times m}{\boldsymbol{A}^{\top}}$, respectively. But this may be costly to accomplish. Instead, we use the fact that $\boldsymbol{U}$ with column vectors in the co-domain $\mathbb{R}^m$ are the images of the column vectors of $\boldsymbol{V}$ in the domain $\mathbb{R}^n$ under $\boldsymbol{A}$. In other words, we find the right singular vectors via the eigen-decomposition of $\underset{n\times m}{\boldsymbol{A}^{\top}}\,\underset{m\times n}{\boldsymbol{A}}$ first, then we construct the left singular vectors from the images of the right singular vectors under $\boldsymbol{A}$.

1. Not only are the vectors $\boldsymbol{v}_i$ (columns of $P \in \mathbb{R}^{n\times n}$) orthogonal, the images of the these right singular vectors under $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ are also orthogonal. In other words, the inner products between $\boldsymbol{A}\boldsymbol{v}_i$ and $\boldsymbol{A}\boldsymbol{v}_j$ must be 0 for $i \neq j$. For any two orthogonal eigenvectors $\boldsymbol{v}_i, \boldsymbol{v}_j, i \neq j$, it holds that

$$
\begin{aligned}
(\boldsymbol{A}\boldsymbol{v}_i)^{\top}(\boldsymbol{A}\boldsymbol{v}_j) &= \boldsymbol{v}_i^{\top}\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{v}_j) \\
&= \boldsymbol{v}_i^{\top}\left[\left(\boldsymbol{A}^{\top}\boldsymbol{A}\right)\boldsymbol{v}_j\right] \\
&= \boldsymbol{v}_i^{\top}(\lambda_j\boldsymbol{v}_j) \\
&= \lambda_j\boldsymbol{v}_i^{\top}\boldsymbol{v}_j \\
&= \lambda_j\vec{0} \\
&= \vec{0}
\end{aligned}
$$

To repeat, the right singular vectors $\boldsymbol{v}_i$ in the domain $\mathbb{R}^n$ have images $\boldsymbol{A}\boldsymbol{v}_i$ in the co-domain $\mathbb{R}^m$ that are also orthogonal.

2. Next, we need to pick left-singular vectors that are orthonormal based on the images of the right singular vectors $\boldsymbol{v}_i$ under $\boldsymbol{A}$. We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$$

confirming that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\top}$ also have the relationship $\sigma_i^2 = \lambda_i$. Here, we utilize the fact that the images of the right singular vectors $\boldsymbol{v}_i$ under $\boldsymbol{A}$ have lengths that are the singular values of $\boldsymbol{A}$. Recall that the L2-norm of a vector squared is equal to dot product of the vector with itself:

$$\|\boldsymbol{A}\boldsymbol{v}_i\|^2 = (\boldsymbol{A}\boldsymbol{v}_i)\cdot(\boldsymbol{A}\boldsymbol{v}_i) = \boldsymbol{v}_i^T(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i) = \boldsymbol{v}_i^T(\lambda_i\boldsymbol{v}_i) = \lambda_i(\boldsymbol{v}_i\cdot\boldsymbol{v}_i) = \lambda_i(1) = \lambda_i = \sigma_i^2$$

Therefore:

$$\|\boldsymbol{A}\boldsymbol{v}_i\| = \sigma_i$$

3. The eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$ are the right singular vectors $\boldsymbol{v}_i$

4. The normalized images of the right singular vectors under $\boldsymbol{A}$ are the left-singular vectors $\boldsymbol{u}_i$; this is the same as actually finding the eigen-decomposition of $\boldsymbol{A}\boldsymbol{A}^\top$ and the eigenvectors of orthogonal matrix $\boldsymbol{P} \in \mathbb{R}^{m\times m}$

5. The two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$. We can summarize their connection via the singular value equation

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, \quad i = 1, \ldots, r \tag{3}$$

where $r \in [0, \min(m,n)]$ is the rank of $\boldsymbol{A}$. Where

- For $n < m$ (mapping from lower dimensional vector space to higher dimensional, e.g. $\mathbb{R}^2 \longrightarrow \mathbb{R}^3$), $\min(m,n) = n$ and $\boldsymbol{A}$ has at most n linearly independent eigenvectors ($r \leq n$). The equation 3 therefore holds only for $i \leqslant n$ as it says nothing about the $\boldsymbol{u}_i$ for $i > n$. However, we know by construction (the $\boldsymbol{u}_i$ are the normalized images of the orthogonal $\boldsymbol{v}_i$ vectors) that they are orthonormal.
- For $m < n$ (mapping from higher dimensional vector space to lower dimensional, e.g. $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$), $\min(m,n) = m$ and $\boldsymbol{A}$ has at most m linearly independent eigenvectors ($r \leq m$). The equation 3 therefore holds only for $i \leqslant m$. For $i > m$, we have $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$, since when we map vectors from a higher dimensional vector space to a lower dimensional vector space, some of them would be mapped to the zero vector (information loss). This means that the SVD also supplies an orthonormal basis of the kernel (null space) of $\boldsymbol{A}$— the set of nonzero vectors $\boldsymbol{v}_{m+1}, \cdots, \boldsymbol{v}_n$ that satisfy the system $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$ non-trivially.

6. Concatenating the $\boldsymbol{v}_i$ as the columns of $\boldsymbol{V}$ and the $\boldsymbol{u}_i$ as the columns of $\boldsymbol{U}$ yields

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ has the same dimensions as $\boldsymbol{A}$ and a diagonal structure for rows $1, \ldots, r$. Hence, right-multiplying with $\boldsymbol{V}^\top$ yields $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, which is the SVD of $\boldsymbol{A}$.

1.5.4 Non-Zero Eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^\top$

For any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the matrices $\boldsymbol{A}^\top\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^\top$ possess the same nonzero eigenvalues. Let us assume that λ is a nonzero eigenvalue of $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{x}$ is an eigenvector belonging to that nonzero λ. Thus, the eigenvalue equation

$$\left(\boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{x} = \lambda\boldsymbol{x}$$

can be manipulated by left multiplying by $\boldsymbol{A}^\top$ and pulling on the right-hand side the scalar factor λ forward.

$$\begin{aligned}\boldsymbol{A}^\top\left(\boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{x} &= \boldsymbol{A}^\top(\lambda\boldsymbol{x})\\ &= \lambda\left(\boldsymbol{A}^\top\boldsymbol{x}\right)\end{aligned}$$

and we can use matrix multiplication associativity to reorder the left-hand side factors

$$\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\left(\boldsymbol{A}^\top\boldsymbol{x}\right) = \lambda\left(\boldsymbol{A}^\top\boldsymbol{x}\right)$$

We know that $\left(\boldsymbol{A}^\top\boldsymbol{x}\right)$ is an eigenvector vector, so

$$\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\boldsymbol{v} = \lambda\boldsymbol{v}$$

This is the eigenvalue equation for $\boldsymbol{A}^\top\boldsymbol{A}$. Therefore, λ is the same eigenvalue for

$$\left(\boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{x} = \lambda\boldsymbol{x}$$

and

$$\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\boldsymbol{v} = \lambda\boldsymbol{v}$$

1.6 Eigenvalue Decomposition vs. Singular Value Decomposition

For the eigen-decomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ and the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, the core similarities and difference are as follows:

- The SVD always exists for any matrix $\mathbb{R}^{m\times n}$. The eigen-decomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^n$.

- The vectors in the eigen-decomposition matrix $\boldsymbol{P}$ are not necessarily orthogonal, i.e., the change of basis is not a simple rotation and scaling. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they do represent rotations.

- Both the eigen-decomposition and the SVD are compositions of three linear mappings:

  1. Change of basis in the domain
  2. Independent scaling of each new basis vector and mapping from domain to codomain
  3. Change of basis in the codomain

  A key difference between the eigen-decomposition and the SVD is that in the SVD, domain and codomain can be vector spaces of different dimensions.

- In the SVD, the left- and right-singular vector matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are generally not inverse of each other (they perform basis changes in different vector spaces). In the eigen-decomposition, the basis change matrices $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ are inverses of each other.

- In the SVD, the entries in the diagonal matrix $\boldsymbol{\Sigma}$ are all real and nonnegative, which is not generally true for the diagonal matrix in the eigen-decomposition.
- The SVD and the eigen-decomposition are closely related through their projections
  - The left-singular vectors of $\boldsymbol{A}$ are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^\top$
  - The right-singular vectors of $\boldsymbol{A}$ are eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$.
  - The nonzero singular values of $\boldsymbol{A}$ are the square roots of the nonzero eigenvalues of both $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top\boldsymbol{A}$.
- **For symmetric matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, the eigenvalue decomposition and the SVD are one and the same, which follows from the spectral theorem.**

2 Application: Matrix Approximation

2.1 Rank-1 Matrix

We considered the SVD as a way to factorize $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{m\times n}$ into the product of three matrices, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal and $\boldsymbol{\Sigma}$ contains the singular values on its main diagonal. Besides the full SVD factorization, the truncated SVD can be used to represent a matrix A as a sum of simpler (low-rank) matrices $\boldsymbol{A}_i$. We construct a rank-1 matrix $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}$ as

$$\underset{m\times n}{\boldsymbol{A}_i} := \underset{m\times 1}{\boldsymbol{u}_i} \underset{1\times n}{\boldsymbol{v}_i^\top}$$

which is formed by the outer product of the i th orthogonal column vector of $\boldsymbol{U}$ and $\boldsymbol{V}$.

2.2 Sum of Rank-1 Matrices

Definition 2.1. A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. This form of the SVD is shown in section 1.2. The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i\boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\Sigma_{ij}\boldsymbol{u}_i\boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix (non-diagonal entries where $i \neq j$ are zeros). In addition, any terms $i > r$ also vanish because the corresponding singular values are 0.

2.3 Rank-k Approximation

In the equation above, we summed up the r individual rank-1 matrices to obtain a rank-r matrix $\boldsymbol{A}$. If we do no not sum over all matrices $\boldsymbol{A}_i, i = 1, \ldots, r$, but only up to an intermediate value where $k < r$, we obtain a rank- k approximation

$$\widehat{\boldsymbol{A}}(k) := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$

of $\boldsymbol{A}$ with $\mathrm{rk}(\widehat{\boldsymbol{A}}(k)) = k$.

2.3.1 Efficient Storage

For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we would have to store $m \times n$ entries. The rank-k approximation only requires that we store the k singular values and the k left and right singular vectors. For full matrix, we store

$$m \times n$$

numbers. For the rank-k approximation, we store

$$\begin{aligned} &k + (k \times m) + (k \times n) \\ &k(1 + m + n) \end{aligned}$$

numbers. This is because we have k singular values, $k \times m$ entries for the five left singular vectors, and $k \times n$ entries for the five right singular vectors.

Example

For example, given $\boldsymbol{A} \in \mathbb{R}^{1243 \times 977}$ with rank 700, the full matrix has

$$1243 \times 977 = 1,214,411$$

numbers to store. The rank-400 matrix has

$$400(1 + 1243 + 977) = 888,400$$

numbers to store, which is

$$\frac{888,400}{1,214,411} = 73\%$$

of the original required number of storage.

2.4 Spectral Norm of A Matrix

Definition 2.2. (Spectral Norm of a Matrix). For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{0\}$ (non-zero vectors), the spectral norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}.$$

We use the notation of a subscript in the matrix norm (left-hand side), similar to the Euclidean L2-norm for vectors (right-hand side). **The spectral norm determines, at most, how long any vector $\boldsymbol{x}$ can become when multiplied by $\boldsymbol{A}$.**

Remark. The spectral norm of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is its largest singular value σ_1. In other words,

$$\max_{\boldsymbol{x}} \frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_1$$

1. We compute the eigen-decomposition of the symmetric matrix

$$\begin{aligned} \boldsymbol{A}^\top \boldsymbol{A} &= \boldsymbol{PDP}^{-1} \\ &= \boldsymbol{PDP}^\top \end{aligned}$$

   for diagonal $\boldsymbol{D}$ and orthogonal $\boldsymbol{P}$. Since the columns of $\boldsymbol{P}$ are the right singular vectors and an ONB of $\mathbb{R}^n$ (the domain), we can write every $\boldsymbol{y} \in \mathbb{R}^n$ as a linear combination of the eigenvectors $\boldsymbol{p}_i$ so that

$$\boldsymbol{y} = \boldsymbol{Px} = \sum_{i=1}^{n} x_i \boldsymbol{p}_i, \quad \boldsymbol{x}_{\in}\mathbb{R}$$

   where $\boldsymbol{x}$ is the coordinate vector of $\boldsymbol{y}$ with respective to the eigenbasis. Moreover, since the orthogonal matrix $\boldsymbol{P}$ preserves lengths

$$\begin{aligned} \|\boldsymbol{Px}\|^2 = (\boldsymbol{Px})^\top(\boldsymbol{Px}) = \boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{Px} = \boldsymbol{x}^\top \boldsymbol{Ix} = \boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|^2 \\ \|\boldsymbol{Px}\|^2 = \|\boldsymbol{x}\|^2 \end{aligned}$$

   we see that

$$\|\boldsymbol{y}\|^2 = \|\boldsymbol{Px}\|^2 = \|\boldsymbol{x}\|^2 = \sum_{i=1}^{n} x_i^2$$

2. Then, the dot product of a vector with itself is the squared norm, so

$$\begin{aligned} \|\boldsymbol{Ax}\|_2^2 &= (\boldsymbol{Ax}) \cdot (\boldsymbol{Ax}) \\ &= (\boldsymbol{Ax})^\top (\boldsymbol{Ax}) \\ &= \boldsymbol{x}^\top \boldsymbol{A}^\top (\boldsymbol{Ax}) \\ &= \boldsymbol{x}^\top \boldsymbol{PDP}^\top \boldsymbol{x} \\ &= (\boldsymbol{x}^\top \boldsymbol{P}) \boldsymbol{D} (\boldsymbol{P}^\top \boldsymbol{x}) \qquad \text{For orthogonal matrix } \boldsymbol{P} = \boldsymbol{P}^\top \text{ so } (\boldsymbol{P}^\top \boldsymbol{x}) = (\boldsymbol{Px}) \\ &= \boldsymbol{y}^\top \boldsymbol{Dy} \qquad \text{Because } \boldsymbol{y}^\top = (\boldsymbol{Px})^\top = \boldsymbol{x}^\top \boldsymbol{P} \\ &= \boldsymbol{y} \cdot \boldsymbol{Dy} \\ &= (\sum_{i=1}^{n} x_i \boldsymbol{p}_i) \cdot (\sum_{i=1}^{n} \lambda_i x_i \boldsymbol{p}_i) \qquad \text{The diagonal entries of } \boldsymbol{D} \text{ are the eigenvalues } \lambda_i \text{ of } \boldsymbol{A}^\top \boldsymbol{A} \\ &= (\sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i) \cdot (\sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i) \\ &= \left\langle \sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i, \sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i \right\rangle \end{aligned}$$

3. Using the bilinearity of the dot product

$$
\begin{aligned}
\|\boldsymbol{A}\boldsymbol{x}\|_2^2 &= \sum_{i=1}^{n} \lambda_i \langle x_i \boldsymbol{p}_i, x_i \boldsymbol{p}_i \rangle \\
&= \sum_{i=1}^{n} \lambda_i x_i^2
\end{aligned}
$$

   where we exploited that the $\boldsymbol{p}_i$ are an ONB and $\boldsymbol{p}_i \cdot \boldsymbol{p}_i = \boldsymbol{p}_i^\top \boldsymbol{p}_i = 1$.

4. We know that the spectral norm shows how long any vector $\boldsymbol{x}$ can become when multiplied by $\boldsymbol{A}$

$$
\|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leqslant \left(\max_{1 \leqslant j \leqslant n} \lambda_j \right) \sum_{i=1}^{n} x_i^2 = \max_{1 \leqslant j \leqslant n} \lambda_j \|\boldsymbol{x}\|^2
$$

   so that

$$
\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \leqslant \max_{1 \leqslant j \leqslant n} \lambda_j,
$$

   where λ_j are the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$

5. Taking the square root of both sides and assuming the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$ are sorted in descending order, we get

$$
\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \sqrt{\lambda_1} = \sigma_1,
$$

   where σ_1 is the maximum singular value of $\boldsymbol{A}$

2.5 Eckart-Young Theorem

Definition 2.3. (Eckart-Young Theorem (Eckart and Young, 1936)). Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r and let $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ be a matrix of rank k. For any $k \leqslant r$ with $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ it holds that

$$
\begin{aligned}
\widehat{\boldsymbol{A}}(k) &= \operatorname{argmin}_{\operatorname{rk}(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \\
\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 &= \sigma_{k+1}.
\end{aligned}
$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank- k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most- k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.