Singular value decomposition

We know how to express many linear maps in the form of a diagonal map using a change of variable, by diagonalisation. If $\vec{A}$ is $n \times n$ and diagonalisable, we can simply express it as $\vec{A} = \vec{P}\vec{D}\vec{P}^{-1}$, or $\vec{A} = \vec{P}\vec{D}\vec{P}^T$ if $\vec{A}$ is symmetric. However, for most linear maps, $\vec{A}$ is either not square, or square but not diagonalisable.

Square matrices also have the concepts of eigenvalues and inverses. However, those do not exist for non-square matrices. It is easy to see that for any $\vec{A}_{m\times n}$, if $m \neq n$ then there is no vector $\vec{v}$ such that $\vec{A}\vec{v} = \lambda\vec{v}$, as the dimensions are different on either side of the equation. Inverses are simply not defined for non-square matrices.

We can "express" a non-square matrix as a square matrix by multiplying it by its transpose. For example, given $\vec{A}_{m\times n}$, we can create a new, **symmetric** matrix $\vec{B} = \vec{A}^T\vec{A}$ whose dimension is $n \times n$. We know that $\vec{B}$ has n real eigenvalues, but we can show that they are all positive.

Proof. Let $\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_n\}$ be an orthonormal eigenbasis of $\mathbb{R}^n$ corresponding to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, for the matrix $\vec{B} = \vec{A}^T\vec{A}$. It can be seen that $\left\|\vec{A}\vec{v}_i\right\|^2 = (\vec{A}\vec{v}_i)\cdot(\vec{A}\vec{v}_i) = \vec{v}_i \cdot (\vec{A}^T\vec{A}\vec{v}_i) = \vec{v}_i \cdot (\lambda_i\vec{v}_i) = \lambda_i$. The last step comes from the fact that the basis given is orthonormal, and so all vectors have magnitude 1. So, we have that $\lambda = \left\|\vec{A}\vec{v}_i\right\|^2$, which means all eigenvalues are positive. $\square$

We can define the **singular values** of $\vec{A}$ as the square roots of the eigenvalues of $\vec{B}$, that is, $\sigma_i = \sqrt{\lambda_i}$. We know that all the eigenvalues will be positive, so all singular values will be real and positive. It should also be noted that $\sigma_i = \left\|\vec{A}\vec{v}_i\right\|$.

Theorem: Let $\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_n\}$ be an orthonormal eigenbasis of $\mathbb{R}^n$ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ of $\vec{B} = \vec{A}^T\vec{A}$. Assume that r singular values $\vec{A}$ are equal to 0. Then, $\left\{\vec{A}\vec{v}_1, \vec{A}\vec{v}_2 \cdots \vec{A}\vec{v}_r\right\}$ is an orthogonal basis for col $\vec{A}$, and therefore rank $\vec{A} = r$.

Proof. For any $i \neq j$, $(\vec{A}\vec{v}_i)/cdot(\vec{A}\vec{v}_j) = \vec{v}_i \cdot (\vec{A}^T\vec{A}\vec{v}_j) = \vec{v}_i \cdot (\lambda_j\vec{v}_j) = \lambda_j(\vec{v}_i \cdot \vec{v}_j) = 0$. So, we know that all vectors in the set $\left\{\vec{A}\vec{v}_1, \vec{A}\vec{v}_2 \cdots \vec{A}\vec{v}_n\right\}$

are either orthogonal to each other or $\vec{0}$. But we also know that $\left\|\vec{A}\vec{v}_i\right\| = \sigma_i$, so we know which vectors we should remove from this set to retain a set $\left\{\vec{A}\vec{v}_1, \vec{A}\vec{v}_2 \cdots \vec{A}\vec{v}_r\right\}$ which *is* orthogonal. We know that this set is linearly independent as all the vectors are orthogonal, and it can also be shown that its span is equal to col $\vec{A}$. It is given that $\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_n\}$ spans $\mathbb{R}^n$, so $\vec{x} \in \mathbb{R}^n \implies \vec{x} = \sum_{i=1}^{n} c_i \vec{v}_i$ for some constants c_i. Therefore, $\vec{A}\vec{x} = \sum_{i=1}^{n} c_i \vec{A}\vec{v}_i$, that is, any vector in col $\vec{A}$ is also in the span of $\left\{\vec{A}\vec{v}_1, \vec{A}\vec{v}_2 \cdots \vec{A}\vec{v}_r\right\}$. $\square$

Finally, we can define the theoram of single value decomposition. Any matrix $\vec{A}_{m\times n}$ has r nonzero singular values, where $r \leq m, n$. We create a matrix $\vec{\Sigma}_{m\times n}$ with these singular values on the main diagonal, and the rest of the entries equal to 0. Then, there exist two orthogonal matrices $\vec{U}_{m\times m}$ and $\vec{V}_{n\times n}$ such that $\vec{A} = \vec{U}\vec{\Sigma}\vec{V}^T$. These two matrices are not unique, and consist of **singular vectors**: the columns of $\vec{U}$ are **left singular** and the columns of $\vec{V}$ are **right singular**. Most importantly is the remark that connect singular value decomposition to diagonalisation. If $\vec{A}$ is $n \times n$ and diagonalisable, then, if we use an eigenbasis for $\mathbb{R}^n$, in the new coordinate system the map is given by a diagonal matrix ($\vec{D}$). Now, for any $\vec{A}_{m\times n}$ ($\mathbb{R}^n \to \mathbb{R}^m$, if we use as bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ the column spaces of $\vec{V}$ and $\vec{U}$, respectively, then in the new coordinates the map is also diagonal ($\vec{\Sigma}$)!

Proof. Given $\vec{A} : \mathbb{R}^n \to \mathbb{R}^m$, $\vec{y} = \vec{A}\vec{x} = (\vec{U}\vec{\Sigma}\vec{V}^T)\vec{x}$. Then $\vec{U}^T\vec{y} = \vec{\Sigma}(\vec{V}^T\vec{x})$, and $\vec{U}^{-1}\vec{y} = \vec{\Sigma}(\vec{V}^{-1}\vec{x})$. If the columns of $\vec{V}$ and $\vec{U}$ are used as bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we have by definition that $[\vec{y}]_{\vec{U}} = \vec{\Sigma}[\vec{x}]_{\vec{V}}$. $\square$

We can also create an orthonormal basis for col $\vec{A}$. Given an orthogonal basis $\left\{\vec{A}\vec{v}_1, \vec{A}\vec{v}_2 \cdots \vec{A}\vec{v}_r\right\}$, as before, we can create a set $\{\vec{u}_1, \vec{u}_2 \cdots \vec{u}_r\}$ where $\vec{u}_i = \frac{1}{\sigma_i}\vec{A}\vec{v}$; this set will be an orthonormal basis.

Finally, we need to define the earlier referenced notion of singular vectors. Using the set as defined above, we have $\vec{A}\vec{v}_i = \sigma_i \vec{u}_i$. This implies that $\vec{A}^T\vec{A}\vec{v}_i = \sigma_i \vec{A}^T\vec{u}_i$. Since $\vec{v}_i$ is an eigenvector of $\vec{A}^T\vec{A}$, $\vec{A}^T\vec{u}_i = \sigma_i \vec{v}_i \implies \vec{u}_i^T\vec{A} = \sigma_i \vec{v}_i^T$. We define $\vec{v}_i$ as **right singular vectors** and their corresponding $\vec{u}_i$ as **left singular vectors**.

Therefore, we have a way to create the matrices $\vec{V}$ and $\vec{U}$. $\vec{V}$ is easy, since it is filled with the vectors $\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_n\}$ as its columns. The columns of $\vec{U}$ are formed using $\{\vec{u}_1, \vec{u}_2 \cdots \vec{u}_r\}$. However, if $r < m$, then we need to fill up the rest of the matrix with the orthonormal basis for col $\vec{A}$, which we know how to calculate. This guarantees the orthogonal properties of $\vec{V}$ and $\vec{U}$.