The Singular Value Decomposition

We are interested in more than just sym+def matrices. But the eigenvalue decompositions discussed in the last section of notes will play a major role in solving general systems of equations

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \qquad \boldsymbol{y} \in \mathbb{R}^M, \;\; \boldsymbol{A} \text{ is } M \times N \;, \;\; \boldsymbol{x} \in \mathbb{R}^N.$$

We have seen that a symmetric positive definite matrix can be decomposed as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}$, where $\boldsymbol{V}$ is an orthogonal matrix ($\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \mathbf{I}$) whose columns are the eigenvectors of $\boldsymbol{A}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$. Because both orthogonal and diagonal matrices are trivial to invert, this eigenvalue decomposition makes it very easy to solve systems of equations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and analyze the stability of these solutions.

The **singular value decomposition** (SVD) takes apart an arbitrary $M \times N$ matrix $\boldsymbol{A}$ in a similar manner. The SVD of a $M \times N$ matrix $\boldsymbol{A}$ with rank[1] R is

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

where

1. $\boldsymbol{U}$ is a $M \times R$ matrix

   $$\boldsymbol{U} = \left[\boldsymbol{u}_1 \;\middle|\; \boldsymbol{u}_2 \;\middle|\; \cdots \;\middle|\; \boldsymbol{u}_R\right],$$

   whose columns $\boldsymbol{u}_m \in \mathbb{R}^M$ are orthogonal. Note that while $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \mathbf{I}$, in general $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}} \neq \mathbf{I}$ when $R < M$. The columns of $\boldsymbol{U}$ are an orthobasis for the range space of $\boldsymbol{A}$.

[1] Recall that the rank of a matrix is the number of linearly independent columns of a matrix (which is always equal to the number of linearly independent rows).

2. $\boldsymbol{V}$ is a $N \times R$ matrix

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & | & \boldsymbol{v}_2 & | & \cdots & | & \boldsymbol{v}_R \end{bmatrix},$$

whose columns $\boldsymbol{v}_n \in R^N$ are orthonormal. Again, while $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \mathbf{I}$, in general $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} \neq \mathbf{I}$ when $R < N$. The columns of $\boldsymbol{V}$ are an orthobasis for the range space of $\boldsymbol{A}^{\mathrm{T}}$ (recall that Range($\boldsymbol{A}^{\mathrm{T}}$) consists of everything which is orthogonal to the nullspace of $\boldsymbol{A}$).

3. $\boldsymbol{\Sigma}$ is a $R \times R$ diagonal matrix with positive entries:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & \\ 0 & \sigma_2 & 0 & \cdots & \\ \vdots & & \ddots & & \\ 0 & \cdots & \cdots & \sigma_R & \end{bmatrix}.$$

We call the σ_r the **singular values** of $\boldsymbol{A}$. By convention, we will order them such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R$.

4. The $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_R$ are eigenvectors of the positive semi-definite matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. Note that

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathrm{T}},$$

and so the singular values $\sigma_1, \ldots, \sigma_R$ are the square roots of the non-zero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

5. Similarly,

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\mathrm{T}},$$

and so the $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_R$ are eigenvectors of the positive semi-definite matrix $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$. Since the non-zero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$

and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are the same, the σ_r are also square roots of the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.

6. The rank R is the number of linearly independent columns of $\boldsymbol{A}$; this is the same as the number of linearly independent rows. Thus $R \leq \min(M, N)$. We say $\boldsymbol{A}$ is **full rank** if $R = \min(M, N)$.

7. As $\boldsymbol{A}$ is rank R, its rows span an R-dimensional linear subspace of $\mathbb{R}^N$. This is called the **row space** of $\boldsymbol{A}$:
$$\begin{aligned} \text{row space} &= \operatorname{Range}(\boldsymbol{A}^{\mathrm{T}}) \\ &= \{\boldsymbol{w} \in \mathbb{R}^N \ : \ \boldsymbol{w} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z} \text{ for some } \boldsymbol{z} \in \mathbb{R}^M\}. \end{aligned}$$
The columns of $\boldsymbol{V}$ can be interpreted as and orthobasis for this space.

8. The **null space** of $\boldsymbol{A}$,
$$\operatorname{Null}(\boldsymbol{A}) = \{\boldsymbol{w} \in \mathbb{R}^N \ : \ \boldsymbol{A}\boldsymbol{w} = \boldsymbol{0}\},$$
is orthogonal to the row space. For $\boldsymbol{x}_1 \in \operatorname{Range}(\boldsymbol{A}^{\mathrm{T}})$ and $\boldsymbol{x}_2 \in \operatorname{Null}(\boldsymbol{A})$, we have
$$\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{z}, \boldsymbol{0} \rangle = 0.$$
The null space has dimension $N - R$, and so is spanned by some set of orthonormal basis vectors $\boldsymbol{v}_{R+1}, \ldots, \boldsymbol{v}_N$ that we can collect into an $N \times (N - R)$ matrix $\boldsymbol{V}_0$:
$$\boldsymbol{V}_0 = \begin{bmatrix} \boldsymbol{v}_{R+1} & | & \boldsymbol{v}_{R+2} & | & \cdots & | & \boldsymbol{v}_N \end{bmatrix}.$$
Note that $\boldsymbol{V}_0^{\mathrm{T}}\boldsymbol{V}_0 = \mathbf{I}$ and $\boldsymbol{V}_0^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{0}$.

9. As $\boldsymbol{A}$ is rank R, it columns span an R-dimensional subspace of $\mathbb{R}^M$. This is called the **column space** of $\boldsymbol{A}$:

$$\begin{aligned} \text{column space} &= \text{Range}(\boldsymbol{A}) \\ &= \{\boldsymbol{z} \in \mathbb{R}^M \ : \ \boldsymbol{z} = \boldsymbol{A}\boldsymbol{w} \text{ for some } \boldsymbol{w} \in \mathbb{R}^N\}. \end{aligned}$$

   The columns of $\boldsymbol{U}$ can be interpreted as and orthobasis for this space.

10. The null space of $\boldsymbol{A}^{\mathrm{T}}$, sometime referred to as the **left null space** of $\boldsymbol{A}$,

$$\text{Null}(\boldsymbol{A}^{\mathrm{T}}) = \{\boldsymbol{z} \in \mathbb{R}^M \ : \ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{0}\},$$

   is orthogonal to the column space. For $\boldsymbol{y}_1 \in \text{Range}(\boldsymbol{A})$ and $\boldsymbol{y}_2 \in \text{Null}(\boldsymbol{A}^{\mathrm{T}})$, we have

$$\langle \boldsymbol{y}_1, \boldsymbol{y}_2 \rangle = \langle \boldsymbol{A}\boldsymbol{w}, \boldsymbol{y}_2 \rangle = \langle \boldsymbol{w}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}_2 \rangle = \langle \boldsymbol{w}, \boldsymbol{0} \rangle = 0.$$

   The left null space has dimension $M - R$, and so is spanned by some set of orthonormal basis vectors $\boldsymbol{u}_{R+1}, \ldots, \boldsymbol{u}_M$ that we can collect into an $M \times (M - R)$ matrix $\boldsymbol{U}_0$:

$$\boldsymbol{U}_0 = \begin{bmatrix} \boldsymbol{u}_{R+1} & | & \boldsymbol{u}_{R+2} & | & \cdots & | & \boldsymbol{u}_N \end{bmatrix}.$$

   Note that $\boldsymbol{U}_0^{\mathrm{T}}\boldsymbol{U}_0 = \mathbf{I}$ and $\boldsymbol{U}_0^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{0}$.

11. An equivalent way to write the SVD is as

$$\boldsymbol{A} = \boldsymbol{U}_{\text{full}}\boldsymbol{\Sigma}_{\text{full}}\boldsymbol{V}_{\text{full}}^{\mathrm{T}},$$

   where

$$\boldsymbol{U}_{\text{full}} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_0 \end{bmatrix}, \quad \boldsymbol{V}_{\text{full}} = \begin{bmatrix} \boldsymbol{V} & \boldsymbol{V}_0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\text{full}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0}_{R\times(N-R)} \\ \boldsymbol{0}_{(M-R)\times R} & \boldsymbol{0}_{(M-R)\times(N-R)} \end{bmatrix}.$$

Now, $\boldsymbol{U}_{\text{full}}$ is an $M \times M$ orthonormal matrix with $\boldsymbol{U}_{\text{full}}\boldsymbol{U}_{\text{full}}^{\mathrm{T}} = \mathbf{I}$, similarly $\boldsymbol{V}_{\text{full}}$ is $N \times N$ with $\boldsymbol{V}_{\text{full}}\boldsymbol{V}_{\text{full}}^{\mathrm{T}} = \mathbf{I}$, and $\boldsymbol{\Sigma}_{\text{full}}$ is $M \times N$ (the same sizes as $\boldsymbol{A}$) with a diagonal matrix in its upper left corner. In fact, this is the factorization the MATLAB command `svd` returns.

As before, we will often times find it useful to write the SVD as the sum of R rank-1 matrices:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \sum_{r=1}^{R} \sigma_r\, \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}.$$

When $\boldsymbol{A}$ is **overdetermined** ($M > N$), the decomposition looks like this

$$\begin{bmatrix} \\ \\ \boldsymbol{A} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \boldsymbol{U} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_R \end{bmatrix} \begin{bmatrix} & & \\ & \boldsymbol{V}^{\mathrm{T}} & \\ & & \end{bmatrix}$$

When $\boldsymbol{A}$ is **underdetermined** ($M < N$), the SVD looks like this

$$\begin{bmatrix} & & & \\ & \boldsymbol{A} & & \\ & & & \end{bmatrix} = \begin{bmatrix} & & \\ & \boldsymbol{U} & \\ & & \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_R \end{bmatrix} \begin{bmatrix} & & & & \\ & & \boldsymbol{V}^{\mathrm{T}} & & \\ & & & & \end{bmatrix}$$

When $\boldsymbol{A}$ is **square** and full rank ($M = N = R$), the SVD looks like

$$\begin{bmatrix} & & \\ & \boldsymbol{A} & \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ & \boldsymbol{U} & \\ & & \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{bmatrix} \begin{bmatrix} & & \\ & \boldsymbol{V}^{\mathrm{T}} & \\ & & \end{bmatrix}$$

Technical Details: Existence of the SVD

In this section we will prove that any $M \times N$ matrix $\boldsymbol{A}$ with $\operatorname{rank}(\boldsymbol{A}) = R$ can be written as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

where $\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}$ have the five properties listed at the beginning of the last section.

Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric positive semi-definite, we can write:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}},$$

where the $\boldsymbol{v}_n$ are orthonormal and the λ_n are real and non-negative. Since $\operatorname{rank}(\boldsymbol{A}) = R$, we also have $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = R$, and so $\lambda_1, \ldots, \lambda_R$ are all strictly positive above, and $\lambda_{R+1} = \cdots = \lambda_N = 0$.

Set

$$\boldsymbol{u}_m = \frac{1}{\sqrt{\lambda_m}} \boldsymbol{A}\boldsymbol{v}_m, \quad \text{for } m = 1, \ldots, R, \qquad \boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_R \end{bmatrix}.$$

Notice that these $\boldsymbol{u}_m$ are orthonormal, as

$$\langle \boldsymbol{u}_m, \boldsymbol{u}_\ell \rangle = \frac{1}{\sqrt{\lambda_m \lambda_\ell}} \boldsymbol{v}_\ell^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{v}_m = \sqrt{\frac{\lambda_m}{\lambda_\ell}}\, \boldsymbol{v}_\ell^{\mathrm{T}} \boldsymbol{v}_m = \begin{cases} 1, & m = \ell, \\ 0, & m \neq \ell. \end{cases}$$

These $\boldsymbol{u}_m$ also happen to be eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$, as

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_m = \frac{1}{\sqrt{\lambda_m}} \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_m = \sqrt{\lambda_m}\boldsymbol{A}\boldsymbol{v}_m = \lambda_m \boldsymbol{u}_m.$$

Now let $\boldsymbol{u}_{R+1}, \ldots, \boldsymbol{u}_M$ be an orthobasis for the null space of $\boldsymbol{U}^{\mathrm{T}}$ — concatenating these two sets into $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ forms an orthobasis for all of $\mathbb{R}^M$.

Let

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_R\end{bmatrix}, \quad \boldsymbol{V}_0 = \begin{bmatrix}\boldsymbol{v}_{R+1} & \boldsymbol{v}_{R+2} & \cdots & \boldsymbol{v}_N\end{bmatrix}, \quad \boldsymbol{V}_{\text{full}} = \begin{bmatrix}\boldsymbol{V} & \boldsymbol{V}_0\end{bmatrix}$$

and

$$\boldsymbol{U}_0 = \begin{bmatrix}\boldsymbol{u}_{R+1} & \boldsymbol{u}_{R+2} & \cdots & \boldsymbol{u}_M\end{bmatrix}, \quad \boldsymbol{U}_{\text{full}} = \begin{bmatrix}\boldsymbol{U} & \boldsymbol{U}_0\end{bmatrix}.$$

It should be clear that $\boldsymbol{V}_{\text{full}}$ is an $N \times N$ orthonormal matrix and $\boldsymbol{U}_{\text{full}}$ is a $M \times M$ orthonormal matrix. Consider the $M \times N$ matrix $\boldsymbol{U}_{\text{full}}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}_{\text{full}}$ — the entry in the mth rows and nth column of this matrix is

$$(\boldsymbol{U}_{\text{full}}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}_{\text{full}})[m,n] = \boldsymbol{u}_m^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_n = \begin{cases}\sqrt{\lambda_n}\,\boldsymbol{u}_m^{\mathrm{T}}\boldsymbol{u}_n & n = 1,\ldots,R\\ 0, & n = R+1,\ldots,N.\end{cases}$$

$$= \begin{cases}\sqrt{\lambda_n}, & m = n = 1,\ldots,R\\ 0, & \text{otherwise.}\end{cases}$$

Thus

$$\boldsymbol{U}_{\text{full}}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}_{\text{full}} = \boldsymbol{\Sigma}_{\text{full}}$$

where

$$\Sigma_{\text{full}}[m,n] = \begin{cases}\sqrt{\lambda_n}, & m = n = 1,\ldots,R\\ 0, & \text{otherwise.}\end{cases}$$

Since $\boldsymbol{U}_{\text{full}}\boldsymbol{U}_{\text{full}}^{\mathrm{T}} = \mathbf{I}$ and $\boldsymbol{V}_{\text{full}}\boldsymbol{V}_{\text{full}}^{\mathrm{T}} = \mathbf{I}$, we have

$$\boldsymbol{A} = \boldsymbol{U}_{\text{full}}\boldsymbol{\Sigma}_{\text{full}}\boldsymbol{V}_{\text{full}}^{\mathrm{T}}.$$

Since $\boldsymbol{\Sigma}_{\text{full}}$ is non-zero only in the first R locations along its main diagonal, the above reduces to

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}, \quad \boldsymbol{\Sigma} = \begin{bmatrix}\sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_R}\end{bmatrix}.$$