Chapter 2

Singular Value Decomposition

2.1 Reduced and Full Singular Value Decompositions

In the last chapter, we saw that, if $\mathsf{A} \in \mathbb{C}^{n\times n}$ is Hermitian then it is unitarily diagonalizable, i.e., there is a diagonal matrix $\mathsf{D} \in \mathbb{C}^{n\times n}$ and a unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$ for which we have

$$\mathsf{A} = \mathsf{U}^H \mathsf{D} \mathsf{U}.$$

This gives us, at least for this class of matrices, a nice geometric interpretation of the action of a matrix. Namely,

a) Since U is unitary and $\|\mathsf{U}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$, the action of U is *essentially* that of a rotation/reflection (it does not change the magnitude of the vector).

b) The matrix D is diagonal; its action is a dilation in each coordinate direction.

c) Finally, $\mathsf{U}^H = \mathsf{U}^{-1}$ reverses the rotation/reflection and brings us back to the original orientation.

We would like to have an analogue of such a decomposition not only for Hermitian matrices but for every matrix. This is the purpose of the so-called *singular value decomposition* (SVD).

Let $\mathsf{A} \in \mathbb{C}^{m\times n}$. It is a non-trivial geometric fact that the image of the unit ball in $\mathbb{C}^n$ under the action of A is a hyper-ellipse in $\mathbb{C}^m$. To describe a generic hyper-ellipse in $\mathbb{C}^m$, one needs only specify its principal semi-axes. To this end, there is an orthonormal basis $\{\boldsymbol{u}_i\}_{i=1}^m$ of $\mathbb{C}^m$, such that, for some $\sigma_i \geq 0$, the principal semi-axes of this hyper-ellipse are $\{\sigma_i \boldsymbol{u}_i\}_{i=1}^m$. It turns out that for our hyper-ellipse, since $m \geq n$, no more than n of these scalars σ_i can be greater than zero. In other words, the principal semi-axes of our image can be described by a smaller collection: $\{\sigma_i \boldsymbol{u}_i\}_{i=1}^n$.

For now and for simplicity, assume that $m \geq n = \text{rank}(\mathsf{A})$. We define the following objects:

1. The *singular values* are the non-negative scalars $\{\sigma_i\}_{i=1}^n$ introduced above, which are arranged (typically) in non-increasing order. (In the present case, they must all be positive, since $\text{rank}(\mathsf{A}) = n$.)

2. The *left singular vectors* are the orthonormal vectors $\{\boldsymbol{u}_i\}_{i=1}^n \subset \mathbb{C}^m$. (Note that, this collection is not large enough to make a basis for $\mathbb{C}^m$.)

3. The *right singular vectors* $\{\boldsymbol{v}_i\}_{i=1}^n \subset \mathbb{C}^n$ are orthonormal vectors defined by the relation $\mathsf{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$. These vectors represent the principal semi-axes of the unit ball in $\mathbb{C}^n$.

If you believe the story above, then we can describe the action of A via

$$\mathsf{A}[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n] = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}.$$

In other words, we have obtained the representation

$$\mathsf{AV} = \hat{\mathsf{U}}\hat{\Sigma}$$

where

1. $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ is diagonal and has positive entries.

2. $\hat{\mathsf{U}} \in \mathbb{C}^{m \times n}$ has orthonormal columns.

3. $\mathsf{V} \in \mathbb{C}^{n \times n}$ is unitary.

This is the so-called *reduced SVD* of a matrix. If one wants to obtain the *full SVD* we notice that the columns of $\hat{\mathsf{U}}$ can be completed to a full orthonormal basis of $\mathbb{C}^m$ to obtain

$$\mathsf{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n, \boldsymbol{u}_{n+1}, \cdots, \boldsymbol{u}_m] \in \mathbb{C}^{m \times m},$$

which is square and unitary. Next, we define

$$\Sigma = \begin{bmatrix} \hat{\Sigma} \\ \mathsf{O} \end{bmatrix},$$

where $\mathsf{O} \in \mathbb{C}^{(m-n)\times n}$ is a matrix of zeros. It is easy to see that the reduced SVD is equivalent to the representation

$$\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H.$$

This motivates the following definition.

Definition 2.1.1 (SVD)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$. *A* singular value decomposition *(SVD) of the matrix* A *is a factorization of the form*

$$\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H,$$

where $\mathsf{U} \in \mathbb{C}^{m\times m}$ *and* $\mathsf{V} \in \mathbb{C}^{n\times n}$ *are unitary,* $\Sigma \in \mathbb{R}^{m\times n}$ *is diagonal – meaning that* $[\Sigma]_{i,j} = 0$*, for* $i \neq j$ *– and the diagonal entries* $[\Sigma]_{i,i} = \sigma_i$ *are nonnegative and in non-increasing order:* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ *with* $p = \min(m, n)$.

2.2 Existence and Uniqueness of the SVD

Let us now show that *every* matrix has a SVD.

Theorem 2.2.1 (existence of SVD)**.** *Every matrix* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *has a singular value decomposition. The singular values are unique and*

$$\{\sigma_i^2\}_{i=1}^p = \begin{cases} \sigma(\mathsf{A}^H\mathsf{A}) & \text{if} \quad m \geq n \\ \sigma(\mathsf{A}\mathsf{A}^H) & \text{if} \quad m \leq n \end{cases},$$

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

Proof. (Existence of SVD): Let us set $\sigma_1 = \|\mathsf{A}\|_2$. Using a standard argument from Analysis, by compactness, there is $\boldsymbol{v}_1 \in \mathbb{C}^n$ with $\|\boldsymbol{v}_1\|_{\ell^2(\mathbb{C}^n)} = 1$ such that $\|\mathsf{A}\|_2 = \|\mathsf{A}\boldsymbol{v}_1\|_{\ell^2(\mathbb{C}^m)} = \sigma_1$. Define $\boldsymbol{u}_1 = \|\mathsf{A}\|_2^{-1}\mathsf{A}\boldsymbol{v}_1$, then

$$\mathsf{A}\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1, \qquad \|\boldsymbol{u}_1\|_{\ell^2(\mathbb{C}^m)} = 1.$$

Extend $\{\boldsymbol{v}_1\}$ to an orthonormal basis of $\mathbb{C}^n$ and $\{\boldsymbol{u}_1\}$ to an orthonormal basis of $\mathbb{C}^m$, respectively. In doing so, we obtain matrices

$$\mathsf{U}_1 = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m] \in \mathbb{C}^{m\times m}, \quad \mathsf{V}_1 = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n] \in \mathbb{C}^{n\times n},$$

that are unitary and, more importantly, satisfy

$$\mathsf{A}\mathsf{V}_1 = \mathsf{U}_1 \begin{bmatrix} \sigma_1 & \boldsymbol{w}^H \\ \mathsf{O} & \mathsf{B} \end{bmatrix} = \mathsf{U}_1\mathsf{S}$$

for some $\boldsymbol{w} \in \mathbb{C}^{n-1}$ and $\mathsf{B} \in \mathbb{C}^{(m-1)\times(n-1)}$.

Notice that

$$\mathsf{S}\begin{bmatrix}\sigma_1\\ \boldsymbol{w}\end{bmatrix} = \begin{bmatrix}\sigma_1 & \boldsymbol{w}^H\\ \mathsf{O} & \mathsf{B}\end{bmatrix}\begin{bmatrix}\sigma_1\\ \boldsymbol{w}\end{bmatrix} = \begin{bmatrix}\sigma_1^2 + \boldsymbol{w}^H\boldsymbol{w}\\ \mathsf{B}\boldsymbol{w}\end{bmatrix}.$$

Therefore

$$\begin{aligned}\left\|\mathsf{S}\begin{bmatrix}\sigma_1\\ \boldsymbol{w}\end{bmatrix}\right\|_{\ell^2(\mathbb{C}^m)} &= \sqrt{(\sigma_1^2 + \boldsymbol{w}^{\mathrm{H}}\boldsymbol{w})^2 + \|\mathsf{B}\boldsymbol{w}\|^2_{\ell^2(\mathbb{C}^{m-1})}} \geq \sigma_1^2 + \boldsymbol{w}^H\boldsymbol{w}\\ &= \left(\sigma_1^2 + \boldsymbol{w}^H\boldsymbol{w}\right)^{1/2}\left\|\begin{bmatrix}\sigma_1\\ \boldsymbol{w}\end{bmatrix}\right\|_{\ell^2(\mathbb{C}^m)},\end{aligned}$$

which shows the lower bound $\|\mathsf{S}\|_2 \geq \left(\sigma_1^2 + \boldsymbol{w}^H\boldsymbol{w}\right)^{1/2}$. On the other hand, since $\mathsf{U}_1^H\mathsf{A}\mathsf{V}_1 = \mathsf{S}$, and U_1 and V_1 are unitary, then we must have that $\sigma_1 = \|\mathsf{A}\|_2 = \|\mathsf{S}\|_2$, which forces us to have $\boldsymbol{w} = \boldsymbol{0}$.

This shows the result if $n = 1$ or $m = 1$. Otherwise, we proceed by induction upon realizing that B describes the action of the matrix A on $\operatorname{span}(\{\boldsymbol{v}_1\})^\perp$. The induction hypothesis then shows that B has an SVD, say $\mathsf{B} = \mathsf{U}_2\Sigma_2\mathsf{V}_2^H$. Therefore, we have

$$\mathsf{A} = \mathsf{U}_1\begin{bmatrix}1 & \mathsf{O}\\ \mathsf{O} & \mathsf{U}_2\end{bmatrix}\begin{bmatrix}\sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}1 & \mathsf{O}\\ \mathsf{O} & \mathsf{V}_2^H\end{bmatrix}\mathsf{V}_1^H,$$

which is the sought-after SVD for A with

$$\mathsf{U} = \mathsf{U}_1\begin{bmatrix}1 & \mathsf{O}\\ \mathsf{O} & \mathsf{U}_2\end{bmatrix}, \quad \Sigma = \begin{bmatrix}\sigma_1 & \mathsf{O}\\ \mathsf{O} & \Sigma_2\end{bmatrix}, \quad \mathsf{V} = \mathsf{V}_1\begin{bmatrix}1 & \mathsf{O}\\ \mathsf{O} & \mathsf{V}_2\end{bmatrix}.$$

The existence part is finished by induction.

(Uniqueness of Singular Values): Next, notice that

$$\mathsf{A}^H\mathsf{A} = \left(\mathsf{U}\Sigma\mathsf{V}^H\right)^H\mathsf{U}\Sigma\mathsf{V}^H = \mathsf{V}\Sigma\mathsf{U}^H\mathsf{U}\Sigma\mathsf{V}^H = \mathsf{V}\Sigma^\intercal\Sigma\mathsf{V}^H,$$

and

$$\mathsf{A}\mathsf{A}^H = \mathsf{U}\Sigma\mathsf{V}^H\left(\mathsf{U}\Sigma\mathsf{V}^H\right)^H = \mathsf{U}\Sigma\Sigma^\intercal\mathsf{U}^H,$$

where $\Sigma^\intercal\Sigma \in \mathbb{R}^{n\times n}$ and $\Sigma\Sigma^\intercal \in \mathbb{R}^{m\times m}$ are diagonal matrices with the diagonal entries $\sigma_1^2, \cdots, \sigma_p^2$, plus zeros for padding, as needed. Thus $\mathsf{A}^H\mathsf{A} \asymp \Sigma^\intercal\Sigma$ and $\mathsf{A}\mathsf{A}^H \asymp \Sigma\Sigma^\intercal$. Using this fact and the the fact that eigenvalues are uniquely determined proves the result. □

The uniqueness results for the right and left are a little more subtle. We have for instance

Theorem 2.2.2. *Suppose that* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *with* $m \geq n$. *If*

$$\mathsf{A} = \mathsf{U}_1\Sigma_1\mathsf{V}_1^H = \mathsf{U}_2\Sigma_2\mathsf{V}_2^H$$

are two SVDs for A, *then* $\Sigma_1 = \Sigma_2$ *(from the last result), the columns of* V_1 *and* V_2 *form an orthonormal basis of eigenvectors of* $\mathsf{A}^H\mathsf{A}$, *and, if* $\mathsf{A}^H\mathsf{A}$ *has* n *distinct eigenvalues, then*

$$\mathsf{V}_1 = \mathsf{V}_2\mathsf{D},$$

for some $\mathsf{D} = \operatorname{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$, *with angles* $\theta_i \in \mathbb{R}$, $i = 1, \cdots, n$. *Finally, if* $\operatorname{rank}(\mathsf{A}) = n$ *(*A *has full rank), and*

$$\mathsf{A} = \mathsf{U}_1\Sigma\mathsf{V}^H = \mathsf{U}_2\Sigma\mathsf{V}^H$$

are two SVDs for A, *then the first* n *columns of* U_1 *and* U_2 *are equal.*

Proof. Suppose that $\mathsf{A}^H\mathsf{A}$ has distinct eigenvalues and

$$\mathsf{A} = \mathsf{U}_1^{\Sigma}\mathsf{V}_1^H = \mathsf{U}_2^{\Sigma}\mathsf{V}_2^H,$$

where we have used the fact that the singular values are uniquely determined from the last theorem. Also from the last proof,

$$\mathsf{A}^H\mathsf{A} = \mathsf{V}_1\Sigma^\intercal\Sigma\mathsf{V}_1^H = \mathsf{V}_2\Sigma^\intercal\Sigma\mathsf{V}_2^H,$$

which proves that the columns of V_1 and V_2 are eigenvectors of $\mathsf{A}^H\mathsf{A}$.

Now, if the eigenvalues $\lambda_i = \sigma_i^2$ of $\mathsf{A}^H\mathsf{A}$ are all simple, the corresponding eigenspaces $E(\lambda_i, \mathsf{A}^H\mathsf{A})$ are all one-dimensional and

$$E(\lambda_i, \mathsf{A}^H\mathsf{A}) = \operatorname{span}(\{\boldsymbol{v}_{1,i}\}) = \operatorname{span}(\{\boldsymbol{v}_{2,i}\}), \quad i = 1, \cdots, n,$$

with

$$\|\boldsymbol{v}_{1,i}\|_{\ell^2(\mathbb{C}^n)} = \|\boldsymbol{v}_{2,i}\|_{\ell^2(\mathbb{C}^n)} = 1, \quad i = 1, \cdots, n,$$

where $\mathsf{V}_k = [\boldsymbol{v}_{k,1}, \cdots, \boldsymbol{v}_{k,n}]$, $k = 1, 2$. The only possibility is that $\boldsymbol{v}_{1,i} = \gamma_i\boldsymbol{v}_{2,i}$, where $|\gamma_i| = 1$, for $i = 1, \cdots, n$. This proves that $\mathsf{V}_1 = \mathsf{V}_2\mathsf{D}$, where D is a diagonal matrix with the required structure.

Finally, if

$$\mathsf{A} = \mathsf{U}_1\Sigma\mathsf{V}^H = \mathsf{U}_2\Sigma\mathsf{V}^H$$

are two SVDs for A, then we have the following family of equations

$$\mathsf{A}\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_{k,i}, \quad k = 1, 2, \quad i = 1, \cdots, n,$$

where $\mathsf{U}_k = [\boldsymbol{u}_{k,1}, \cdots, \boldsymbol{u}_{k,m}]$, $k = 1, 2$. Since A has full rank and $\sigma_i > 0$, $\boldsymbol{u}_{1,i} = \boldsymbol{u}_{2,i}$, $i = 1, \cdots, n$, is the only possibility. $\square$

Remark 2.2.3 (geometric interpretation of the SVD)**.** *Let $\boldsymbol{b} \in \mathbb{C}^m$ and expand it in the basis of left singular vectors* U. *This gives the coordinate vector $\boldsymbol{b}' = \mathsf{U}^H \boldsymbol{b}$. Do the same for $\boldsymbol{x} \in \mathbb{C}^n$ to obtain it coordinate vector $\boldsymbol{x}' = \mathsf{V}^H \boldsymbol{x}$. Once we have this, notice that, if $\boldsymbol{b} = \mathsf{A}\boldsymbol{x}$ we can proceed as follows*

$$\boldsymbol{b}' = \mathsf{U}^H \boldsymbol{b} = \mathsf{U}^H \mathsf{A}\boldsymbol{x} = \mathsf{U}^H \mathsf{U}\Sigma \mathsf{V}^H \boldsymbol{x} = \Sigma \mathsf{V}^H \boldsymbol{x} = \Sigma \boldsymbol{x}'.$$

In other words, the SVD is essentially saying that every matrix, once proper bases for the domain and target spaces are chosen, may be viewed as a diagonal matrix.

2.3 Properties of the SVD

The main motivation for the SVD was to try to construct an analogue of the spectral decomposition, which we know is only valid for square, nondefective matrices. Let us study now the relation between these two constructions, which in principle are not related to each other.

Theorem 2.3.1 (SVD and rank)**.** *Let $\mathsf{A} \in \mathbb{C}^{m\times n}$, then* $\operatorname{rank}(\mathsf{A})$ *coincides with the number of nonzero singular values.*

Proof. We write the SVD: $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$. Since U and V are unitary they are full rank. Therefore, $\operatorname{rank}(\mathsf{A}) = \operatorname{rank}(\Sigma)$ and since Σ is diagonal the assertion follows. ☐

Theorem 2.3.2. *Let $\mathsf{A} \in \mathbb{C}^{m\times n}$ with $\operatorname{rank}(\mathsf{A}) = r$. Suppose an SVD for* A *is given by $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$, where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ denote the columns of* U *and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ denote the columns of* V*. Then*

$$\langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle = \mathcal{R}(\mathsf{A}) \quad \textit{and} \quad \langle \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n \rangle = \ker(\mathsf{A}).$$

Proof. ($\subseteq$) From the SVD one can easily write $\mathsf{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$, for $i = 1, \ldots, r$. This proves immediately that $\boldsymbol{u}_i \in \mathcal{R}(\mathsf{A})$, for $i = 1, \ldots, r$. Since $\mathcal{R}(\mathsf{A})$ is a vector subspace of $\mathbb{C}^m$, any linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ is in $\mathcal{R}(\mathsf{A})$. Hence $\langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle \subseteq \mathcal{R}(\mathsf{A})$.

($\supseteq$) Let $\boldsymbol{y} \in \mathcal{R}(\mathsf{A})$. Then $\exists\ \boldsymbol{x} \in \mathbb{C}^n$ such that $\mathsf{A}\boldsymbol{x} = \boldsymbol{y}$. This implies that $\mathsf{U}\Sigma\mathsf{V}^H\boldsymbol{x} = \boldsymbol{y}$ for some $\boldsymbol{x}$. Let $\boldsymbol{x}' = \mathsf{V}^H\boldsymbol{x}$. Then for some $\boldsymbol{x}' \in \mathbb{C}^n$, $\mathsf{U}\Sigma\boldsymbol{x}' = \boldsymbol{y}$. Set $\boldsymbol{x}'' = \Sigma\boldsymbol{x}'$. Note that $\boldsymbol{x}'' \in \mathbb{C}^m$ and $x''_{r+1}, \ldots, x''_m = 0$. Hence for some $\boldsymbol{x}'' \in \mathbb{C}^m$, $\mathsf{U}\boldsymbol{x}'' = \boldsymbol{y}$. Now we write

$$\boldsymbol{y} = \mathsf{U}\boldsymbol{x}'' = \sum_{j=1}^{m} x''_j \boldsymbol{u}_j = \sum_{j=1}^{r} x''_j \boldsymbol{u}_j \in \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle\ .$$

This proves that $\langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle \supseteq \mathcal{R}(\mathsf{A})$, and we are done.

($\subseteq$) From the SVD one can easily write $\mathsf{A}\boldsymbol{v}_i = \boldsymbol{0}$, for $i = r+1, \ldots, n$. This proves immediately that $\boldsymbol{v}_i \in \ker(\mathsf{A})$, for $i = r+1, \ldots, n$. Since $\ker(\mathsf{A})$ is vector subspace of $\mathbb{C}^n$, any linear combination of $\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$ is in $\ker(\mathsf{A})$. Hence $\langle \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n \rangle \subseteq \ker(\mathsf{A})$.

($\supseteq$) Let $\boldsymbol{x} \in \ker(\mathsf{A})$. Then $\mathsf{A}\boldsymbol{x} = \boldsymbol{0}$. This implies that $\mathsf{U}\Sigma\mathsf{V}^H\boldsymbol{x} = \boldsymbol{0}$. Let $\boldsymbol{x}' = \mathsf{V}^H\boldsymbol{x}$. This implies that $\boldsymbol{x} = \mathsf{V}\boldsymbol{x}'$ and $\mathsf{U}\Sigma\boldsymbol{x}' = \boldsymbol{0}$. Since U is invertible, this implies that $\Sigma\boldsymbol{x}' = \boldsymbol{0}$. This homogeneous system always has a solution of the form

$$\boldsymbol{x}' = \begin{bmatrix} x_1' = 0 \\ \vdots \\ x_r' = 0 \\ x_{r+1}' = \alpha_{r+1} \\ \vdots \\ x_n' = \alpha_n \end{bmatrix} ,$$

where $\alpha_{r+1}, \ldots, \alpha_n$ are arbitrary. But this shows that

$$\boldsymbol{x} = \mathsf{V}\boldsymbol{x}' = \sum_{j=1}^{n} x_j' \boldsymbol{v}_j = \sum_{j=r+1}^{n} \alpha_j \boldsymbol{v}_j \in \langle \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n \rangle .$$

This proves that $\langle \boldsymbol{v}_{r+1}, \ldots \boldsymbol{v}_n \rangle \supseteq \ker(\mathsf{A})$, and we are done. $\square$

Theorem 2.3.3. *Suppose* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *with* $m \geq n$. $\mathsf{A}^H\mathsf{A}$ *is nonsingular iff* $\operatorname{rank}(\mathsf{A}) = n$.

Proof. Let $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$ be an SVD for A. Then $\mathsf{A}^H\mathsf{A} = \mathsf{V}\Sigma^\intercal\Sigma\mathsf{V}^H$ yields a unitary diagonalization of $\mathsf{A}^H\mathsf{A}$. Note that

$$\Sigma^\intercal\Sigma = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right) ,$$

where $\sigma_1, \ldots, \sigma_n$ are the singular values of A, some of which may be zero.

It is clear that $\operatorname{rank}(\mathsf{A}) = r$, where r is the number of nonzero singular values. Of course, it must be that $r \leq n$. Likewise, $\operatorname{rank}(\mathsf{A}^H\mathsf{A})$ is the number of nonzero elements on the diagonal of $\Sigma^H\Sigma$. This number must also be r. In other words,

$$\operatorname{rank}(\mathsf{A}) = r = \operatorname{rank}(\mathsf{A}^H\mathsf{A}),$$

which proves the result. $\square$

Theorem 2.3.4 (SVD and norms)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$*, then* $\|\mathsf{A}\|_2 = \sigma_1$ *and* $\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2$.

Proof. The first statement is by construction. The second follows from Exercise 1.3.27 and the fact that U and V are unitary. In this case, $\|\mathsf{U}\Sigma\mathsf{V}^H\|_F = \|\Sigma\mathsf{V}^H\|_F = \|\Sigma\|_F$. □

Theorem 2.3.5 (SVD and self-adjoint matrices)**.** *If* $\mathsf{A}^H = \mathsf{A}$ *then the singular values of* A *are the absolute value of its eigenvalues.*

Proof. Since A is self-adjoint, then it is orthogonally diagonalizable and $\sigma(\mathsf{A}) \subset \mathbb{R}$, i.e.,

$$\mathsf{A} = \mathsf{Q}\Lambda\mathsf{Q}^H = \mathsf{Q}|\Lambda|\operatorname{sgn}(\Lambda)\mathsf{Q}^H,$$

where the notation has the obvious meaning. Define $\mathsf{U} = \mathsf{Q}$ and $\mathsf{V}^H = \operatorname{sgn}(\Lambda)\mathsf{Q}^H$ to obtain the SVD of A. Conclude by uniqueness. □

Theorem 2.3.6 (SVD and determinants)**.** *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *(a square matrix), then*

$$|\det(\mathsf{A})| = \prod_{i=1}^{n} \sigma_i.$$

Proof. By the usual rules for determinants

$$|\det(\mathsf{A})| = |\det(\mathsf{U}\Sigma\mathsf{V}^H)| = |\det(\mathsf{U})||\det(\Sigma)||\det(\mathsf{V}^H)| = |\det(\Sigma)| = \prod_{i=1}^{n} \sigma_i,$$

where we have used the facts that (a) the determinant of a unitary matrix is ± 1, (2) the determinant of diagonal matrix is the product of its diagonals, and (3) Σ is diagonal with nonnegative entries. □

2.4 Low rank approximations

Given $\boldsymbol{u} \in \mathbb{C}^m$, $\boldsymbol{v} \in \mathbb{C}^n$ and $\sigma \in \mathbb{C}$, we can define a matrix $\mathsf{A} \in \mathbb{C}^{m\times n}$ via

$$\mathsf{A} = \sigma \boldsymbol{u}\boldsymbol{v}^H =: \sigma \boldsymbol{u} \otimes \boldsymbol{v},$$

which for every $\boldsymbol{x} \in \mathbb{C}^n$ acts as $\mathsf{A}\boldsymbol{x} = \sigma(\boldsymbol{x}, \boldsymbol{v})\boldsymbol{u} \in \operatorname{span}\{\boldsymbol{u}\}$. This shows, as a consequence of Corollary 1.3.11, that the rank of this matrix is exactly equal to one. The question we want to address now is whether *every* matrix can be represented (or at least approximated) by linear combinations of matrices of this form.

Theorem 2.4.1 (low rank representation)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *be such that* $r = \operatorname{rank}(\mathsf{A})$ *and* $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$ *be a SVD. Then*

$$\mathsf{A} = \sum_{j=1}^{r} \sigma_j \boldsymbol{u}_j \otimes \boldsymbol{v}_j, \tag{2.4.1}$$

Proof. It suffices to write Σ as the sum of matrices of the form

$$\Sigma_j = \operatorname{diag}(0, \cdots, 0, \sigma_j, 0, \cdots, 0)$$

where the element σ_j is in the j-th entry. The rest of the details are left for an exercise. □

Let us now prove the so-called Eckhard Young low rank approximation theorem, which states that truncating the SVD of a matrix gives, in a sense, *the best* low rank approximation to it.

Theorem 2.4.2 (Eckhard Young)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *be such that* $r = \operatorname{rank}(\mathsf{A})$*. For* $k < r$ *define* $\mathsf{A}_k = \sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \otimes \boldsymbol{v}_j$*. Denote by* $\mathcal{C}_k$ *the collection of all matrices* $\mathsf{B} \in \mathbb{C}^{m\times n}$ *such that* $\operatorname{rank}(\mathsf{B}) \leq k$*. Then*

$$\|\mathsf{A} - \mathsf{A}_k\|_2 = \sigma_{k+1} = \inf_{\mathsf{B}\in\mathcal{C}_k} \|\mathsf{A} - \mathsf{B}\|_2.$$

and

$$\|\mathsf{A} - \mathsf{A}_k\|_F = \sqrt{\sum_{j=k+1}^{r} \sigma_j^2} = \inf_{\mathsf{B}\in\mathcal{C}_k} \|\mathsf{A} - \mathsf{B}\|_F.$$

Proof. The values of the norms of the difference $\mathsf{A} - \mathsf{A}_k$ follow from the representation (2.4.1) and Theorem 2.3.4. Next, it is straightforward to see that $\operatorname{rank}(\mathsf{A}_k) = k$.

To show that first infimum is attained at A_k, we use a contradiction argument. Namely, let us assume that there is a matrix $\mathsf{B} \in \mathcal{C}_k$ such that

$$\|\mathsf{A} - \mathsf{B}\|_2 < \|\mathsf{A} - \mathsf{A}_k\|_2 = \sigma_{k+1}.$$

Since $\operatorname{rank}(\mathsf{B}) \leq k$, by Theorem 1.3.13, $\ker(\mathsf{B}) \geq n - k$. There is a subspace $W_1 \leq \ker(\mathsf{B})$ such that $\dim(W_1) = n - k$ and, for every $\boldsymbol{w} \in W_1$, $\mathsf{B}\boldsymbol{w} = \mathbf{0}$.

Then, for any $\boldsymbol{w} \in W$, then we must have $\mathsf{A}\boldsymbol{w} = (\mathsf{A} - \mathsf{B})\boldsymbol{w}$ so that

$$\|\mathsf{A}\boldsymbol{w}\|_2 = \|(\mathsf{A} - \mathsf{B})\boldsymbol{w}\|_2 \leq \|\mathsf{A} - \mathsf{B}\|_2 \|\boldsymbol{w}\|_2 < \sigma_{k+1}\|\boldsymbol{w}\|_2.$$

Next, define $W_2 := \operatorname{span}(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k+1}\})$. We claim that, for every $\boldsymbol{w} \in W_2$, $\|\mathsf{A}\boldsymbol{w}\|_2 \geq \sigma_{k+1}\|\boldsymbol{w}\|_2$. To see this, observe that $\boldsymbol{w} \in W_2$ can be written as

$\boldsymbol{w} = \sum_{i=1}^{k+1} \beta_i \boldsymbol{v}_i$, for some $\beta_1, \cdots, \beta_{k+1} \in \mathbb{C}$. By orthonormality, it follows that

$$\|\boldsymbol{w}\|_2 = \sqrt{\sum_{i=1}^{k+1} |\beta_i|^2}.$$

Using orthonormality again, we see that

$$\|\mathsf{A}\boldsymbol{w}\|_2 = \left\| \sum_{i=1}^{k+1} \sigma_i \boldsymbol{u}_i \right\|_2 = \sqrt{\sum_{i=1}^{k+1} \sigma_i^2 |\beta_i|^2} \geq \sigma_{k+1} \sqrt{\sum_{i=1}^{k+1} |\beta_i|^2} = \sigma_{k+1} \|\boldsymbol{w}\|_2 .$$

Thus $W_1 \leq \mathbb{C}^n$ and $W_2 \leq \mathbb{C}^n$, and the sum of the dimensions of these subspaces exceeds n. Therefore, there must be a non-zero vector in their intersection. But this yields a contradiction.

□

2.5 Problems

1. Let $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$ be unitarily equivalent. Prove that they have the same singular values.

2. Show that if A is real, then it has a real SVD ($\mathsf{U} \in \mathbb{R}^{m\times m}$ and $\mathsf{V} \in \mathbb{R}^{n\times n}$).

3. Show that any matrix in $\mathbb{C}^{m\times n}$ is the limit of a sequence of matrices of full rank.

4. Suppose $\mathsf{A} \in \mathbb{C}^{m\times m}$ has an SVD $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$. Find an eigenvalue decomposition of the matrix
$$\begin{bmatrix} \mathsf{O} & \mathsf{A}^H \\ \mathsf{A} & \mathsf{O} \end{bmatrix}$$

5. Let $\mathsf{A} \in \mathbb{C}^{m\times n}$ with rank(A) $= r$. If A has the SVD $\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H$, the *Moore-Penrose pseudoinverse* of A is defined by
$$\mathsf{A}^\dagger = \mathsf{V}\Sigma^\dagger\mathsf{U}^H,$$
where $\Sigma^\dagger = \mathrm{diag}\{\sigma_1^{-1}, \cdots, \sigma_r^{-1}, 0, \cdots, 0\} \in \mathbb{R}^{n\times m}$. Show the following:
   (a) If A is square and A^{-1} exists, then $\mathsf{A}^\dagger = \mathsf{A}^{-1}$.
   (b) If $m \geq n$ and A has full rank, then $\mathsf{A}^\dagger = (\mathsf{A}^H\mathsf{A})^{-1}\mathsf{A}^H$.
   (c) $\mathsf{A}\mathsf{A}^\dagger\mathsf{A} = \mathsf{A}$.
   (d) $\mathsf{A}^\dagger\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger$.