Chapter 8

Singular Value Decomposition

8.1 Diagonalization of Hermitian Matrices

Lemma 8.1.1 (Schur Decomposition)**.** *For any square matrix A, there exists a unitary matrix U such that*

$$U^H AU = T$$

where T is upper triangular. That is, every square matrix is similar to an upper-triangular matrix.

Proof. We prove this lemma by induction on the size n of the matrix. Since it is clearly true for scalars (i.e., matrices of size $n = 1$), the base case is trivial. Now, suppose that the result holds for all $k = 1, 2, \ldots, n - 1$ and let $A \in \mathbb{C}^{n \times n}$. Since every matrix has at least one eigenvector, we let $\underline{u}$ be an eigenvector of A normalized so that $\|\underline{u}\|_2 = 1$. Using the Gram-Schmidt procedure, it is possible to construct an orthonormal basis $\mathcal{B} = \underline{x}_1, \ldots, \underline{x}_n$ for $\mathbb{C}^n$, with $\underline{x}_1 = \underline{u}$. Define the matrix U_n by

$$U_n = \left[\begin{array}{ccc} \underline{x}_1 & \cdots & \underline{x}_n \end{array} \right].$$

Since $\mathcal{B}$ is a basis for $\mathbb{C}^n$, every column of the matrix AU_n can be expressed as a linear combination of vectors in $\mathcal{B}$, say,

$$A\underline{x}_i = \sum_{j=1}^{n} s_{j,i}\underline{x}_j \quad i = 1, \ldots, n.$$

Note that $A\underline{x}_1 = \lambda_1\underline{x}_1$ for some λ_1 since $\underline{x}_1 = \underline{u}$, an eigenvector of A. We can then write

$$AU_n = \left[\begin{array}{ccc} A\underline{x}_1 & \cdots & A\underline{x}_n \end{array} \right] = U_n \left[\begin{array}{cccc} \lambda_1 & s_{1,2} & \ldots & s_{1,n} \\ 0 & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & s_{n,2} & \cdots & s_{n,n} \end{array} \right] = U_n \left[\begin{array}{cc} \lambda_1 & \underline{s}^T \\ \underline{0} & A_{n-1} \end{array} \right],$$

where we have used the convenient notation

$$A_{n-1} = \left[\begin{array}{ccc} s_{2,2} & \cdots & \underline{s}_{2,n} \\ \vdots & \ddots & \vdots \\ s_{n,2} & \cdots & s_{n,n} \end{array} \right]$$

and $\underline{s}^T = (s_{1,2}, \ldots, s_{1,n})$. By the inductive hypothesis, we can write $A_{n-1} = U_{n-1}T_{n-1}U_{n-1}^H$ where T_{n-1} is upper triangular and U_{n-1} is unitary. It follows that

$$\begin{aligned} AU_n = U_n \left[\begin{array}{cc} \lambda_1 & \underline{s}^T \\ \underline{0} & A_{n-1} \end{array} \right] &= U_n \left[\begin{array}{cc} \lambda_1 & \underline{s}^T \\ \underline{0} & U_{n-1}T_{n-1}U_{n-1}^H \end{array} \right] \\ &= U_n \left[\begin{array}{cc} 1 & \underline{0}^T \\ \underline{0} & U_{n-1} \end{array} \right] \left[\begin{array}{cc} \lambda_1 & \underline{s}^T U_{n-1} \\ \underline{0} & T_{n-1} \end{array} \right] \left[\begin{array}{cc} 1 & \underline{0}^T \\ \underline{0} & U_{n-1}^H \end{array} \right]. \end{aligned}$$

Let U be the matrix given by

$$U = U_n \left[\begin{array}{cc} 1 & \underline{0}^T \\ \underline{0} & U_{n-1} \end{array} \right],$$

and note that U is unitary. It follows that

$$U^H AU = \left[\begin{array}{cc} \lambda_1 & \underline{s}^T U_{n-1} \\ \underline{0} & T_{n-1} \end{array} \right].$$

That is, U is a unitary matrix such that $U^H AU$ is upper-triangular. □

We use this lemma to prove the following theorem.

Theorem 8.1.2. *Every Hermitian $n \times n$ matrix A can be diagonalized by a unitary matrix,*

$$U^H AU = \Lambda,$$

where U is unitary and Λ is a diagonal matrix.

Proof. Note that $A^H = A$ and $T = U^H AU$. Consider the matrix T^H given by

$$T^H = (U^H AU)^H = U^H A^H U = U^H AU = T.$$

That is, T is also Hermitian. Since T is upper triangular, this implies that T is a diagonal matrix. We must conclude that every Hermitian matrix is diagonalized by a unitary matrix. □

This proves every Hermitian matrix has a complete set of orthonormal eigenvectors.

8.2 Singular Value Decomposition

The singular value decomposition (SVD) provides a matrix factorization related to the eigenvalue decomposition that works for all matrices. In general, any matrix $A \in \mathbb{C}^{m\times n}$ can be factored into a product of unitary matrices and a diagonal matrix, as explained below.

Theorem 8.2.1. *Let A be a matrix in $\mathbb{C}^{m\times n}$. Then A can be factored as*

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

where $U \in \mathbb{C}^{m\times m}$ is unitary, $V \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma \in \mathbb{R}^{m\times n}$ has the form

$$\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p),$$

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

The diagonal elements of Σ are called the *singular values* of A and are typically ordered so that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0.$$

Proof. Let

$$A^H AV = V \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

be the spectral decomposition of $A^H A$, where the columns of V are orthonormal eigenvectors

$$V = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix},$$

with $\lambda_1, \lambda_2, \cdots, \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$, where $r \leq p$. For $i \leq r$, let

$$\underline{u}_i = \frac{A\underline{v}_i}{\sqrt{\lambda_i}},$$

and observe that

$$\langle \underline{u}_i | \underline{u}_j \rangle = \frac{\underline{v}_j^H A^H A \underline{v}_i}{\sqrt{\lambda_i \lambda_j}} = \frac{\underline{v}_j^H \underline{v}_i \lambda_i}{\sqrt{\lambda_i \lambda_j}} = \delta_{ij}.$$

Also note that $\{\underline{u}_i\}$ are eigenvectors of AA^H since

$$AA^H \underline{u}_i = AA^H A \frac{\underline{v}_i}{\sqrt{\lambda_i}} = \sqrt{\lambda_i} A \underline{v}_i = \lambda_i \underline{u}_i.$$

The set $\{\underline{u}_i : i = 1, \ldots, r\}$ can be extended using the Gram-Schmidt procedure to form an orthonormal basis for $\mathbb{C}^m$. Let

$$U = \begin{bmatrix} \underline{u}_1 & \cdots & \underline{u}_m \end{bmatrix}.$$

For the zero eigenvalues, the eigenvectors must come from the nullspace of AA^H since the eigenvectors with zero eigenvalues are, by construction, orthogonal to the eigenvectors with nonzero eigenvalues that are in the range of AA^H.

For $\underline{u}_i$ where $i \leq r$, we get

$$\underline{u}_i^H AV = \frac{1}{\sqrt{\lambda_i}} \underline{v}_i^H A^H AV = \sqrt{\lambda_i} \underline{e}_i^H.$$

On the other hand, if $i > r$ then $\underline{u}_i^H AV = \underline{0}$. Hence,

$$U^H AV = \text{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}\right) = \Sigma,$$

as desired. $\square$

This proof gives a recipe for computing the SVD of an arbitrary matrix. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 5 & -1 \\ -1 & 5 \end{bmatrix}.$$

The eigenvalue decomposition of $A^H A$ is given by

$$A^H A = \begin{bmatrix} 27 & -9 \\ -9 & 27 \end{bmatrix} = V \Lambda V^H = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 18 & 0 \\ 0 & 36 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right).$$

This implies that $\Sigma_1 = \Lambda^{1/2}$ and $V_1 = V$. Therefore, we can compute $U_1 = AV_1\Sigma_1^{-1}$ with

$$U_1 = \begin{bmatrix} 1 & 1 \\ 5 & -1 \\ -1 & 5 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} \frac{1}{\sqrt{18}} & 0 \\ 0 & \frac{1}{\sqrt{36}} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Putting this all together, we have the compressed SVD

$$A = U_1\Sigma_1 V_1 = \begin{bmatrix} \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{18} & 0 \\ 0 & \sqrt{36} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right).$$

8.3 Properties of the SVD

Many of the important properties of the SVD can be understood better by separating the non-zero singular values from the zero singular values. To do this, we note that every rank r matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition

$$A = U\Sigma V^H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix} = U_1\Sigma_1 V_1^H,$$

where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary and $U_1 \in \mathbb{C}^{m\times r}$, $U_2 \in \mathbb{C}^{m\times m-r}$, $V_1 \in \mathbb{C}^{n\times r}$, and $V_2 \in \mathbb{C}^{n\times n-r}$ have orthonormal columns. The diagonal matrix $\Sigma_1 \in \mathbb{R}^{r\times r}$ contains the non-zero singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

The factorization $A = U\Sigma V^H$ is called the **full SVD** of the matrix A while the factorization $A = U_1\Sigma_1 V_1$ is called the **compact SVD** of A. The compact SVD of a rank-r matrix retains only the r columns of U, V associated with non-zero singular values.

Let X, Y be inner product spaces and let A define a mapping from X to Y. Then, the columns of V_1 form an orthonormal basis for the vectors in X that are mapped to non-zero vectors (i.e., $\mathcal{N}(A)^\perp$) while the columns of V_2 form an orthonormal basis of $\mathcal{N}(A)$. Likewise, the columns of U_1 form a orthonormal basis for the vectors in Y that lie in the range of A while the vectors in U_2 form orthonormal basis for $\mathcal{R}(A)^\perp$. It follows that the full SVD computes orthonormal bases for all of the four fundamental subspaces of the matrix A. For example, it is easy to show that

$$\begin{aligned}\mathcal{R}(A) &= \text{span}(U_1)\\ \mathcal{R}(A^H) &= \text{span}(V_1)\\ \mathcal{N}(A) &= \text{span}(V_2)\\ \mathcal{N}(A^H) &= \text{span}(U_2)\end{aligned}$$

To see this, notice that $A\sum_{i=1}^{t} c_i \underline{v}_i = \sum_{i=1}^{t} c_i \sigma_i \underline{u}_i$.

From this, we can compute easily any projection onto a fundamental subspace. First, we point out that the projection onto the column space of any matrix $W \in \mathbb{C}^{m\times n}$ with orthonormal columns (i.e., $W^H W = I$) is given by

$$P_W = W(W^H W)^{-1} W^H = W W^H.$$

Therefore, the projection matrices for the fundamental subspaces are given by

$$\begin{aligned}P_{\mathcal{R}(A)} &= U_1 U_1^H\\ P_{\mathcal{R}(A^H)} &= V_1 V_1^H\\ P_{\mathcal{N}(A)} &= V_2 V_2^H\\ P_{\mathcal{N}(A^H)} &= U_2 U_2^H\end{aligned}$$

This decomposition also provides a rank revealing decomposition of a rank-r matrix

$$A = \sum_{i=1}^{r} \sigma_i \underline{u}_i \underline{v}_i^H,$$

where $\underline{u}_i$ is the ith column of U and $\underline{v}$ is the ith column of V. This shows A as the sum of r rank-1 matrices. It also allows one to compute

$$\begin{aligned}\|A\|_F &= \sum_{i=1}^{r} \sigma_i^2\\ \|A\|_2 &= \sigma_1\end{aligned}$$

The pseudoinverse of A is also very easy to compute from the SVD. In particular, one finds that

$$A^\dagger = V\Sigma^\dagger U^H = V_1 \Sigma_1^{-1} U_1^H.$$

One can verify this by computing $A^\dagger A$ and $AA^\dagger$. It also follows from the fact that the pseudoinverse of a scalar σ is σ^{-1} if $\sigma \neq 0$ and zero otherwise.