The topic of this lecture is Singular Value Decomposition (SVD). We start by presenting the singular value decomposition of a given matrix A. We then demonstrate how to use it to approximate a given matrix by lower rank matrix.

1 Singular Value Decomposition - SVD

Singular value decomposition may be considered as a substitute for "eigenvalue" decomposition for matrices which are not normal (or not even square). A fundamental theorem in matrix theory states that a normal matrix $N \in \mathbb{M}_n(\mathbb{C})$ is unitarily equivalent to a diagonal matrix, namely $N = U^\star DU$ where $U \in \mathbb{M}_n(\mathbb{C})$ is unitary and $D \in \mathbb{M}_n(\mathbb{C})$ is diagonal. The singular value decomposition theorem states that every matrix $M \in \mathbb{M}_{m\times n}(\mathbb{C})$ with $m \geq n$ can be decomposed to $U^\star DV$ where $U \in M_m(\mathbb{C})$ and $V \in M_n(\mathbb{C})$ are unitary matrices and $D \in \mathbb{M}_{m\times n}(\mathbb{R})$ is non-negative and "diagonal", i.e., D is all zeros except for the singular values $D_{11} = \sigma_1, ..., D_{nn} = \sigma_n$. By convention we will always assume that $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n$There is a close connection between the singular value decomposition of a matrix A and the diagonalization of $AA^\star$ and $A^\star A$.

$$\text{If } A = U^\star DV \text{ then } AA^\star = U^\star D^2 U \text{ and } A^\star A = V^\star D^2 V.$$

We conclude that the singular values of A are the square roots of the non-negative eigenvalues of $AA^\star$ and $A^\star A$, and the rows of the unitary matrices U, V are the eigenvectors of $AA^\star$ and $A^\star A$ respectively. From the singular value decomposition we can see that the rank of A is $\max\{r : \sigma_r > 0\}$. We can also see that the basis for the image and the kernel of the matrix A are first r rows of U and last $m - r$ rows of U respectively.

Theorem 1 *A matrix $A \in \mathbb{M}_{m\times n}(\mathbb{C})$,mgen can be decomposed as $A = UDV^\star$, where $U \in \mathbb{M}_{m\times m}(\mathbb{C})$ and $V \in \mathbb{M}_{n\times n}(\mathbb{C})$ are unitary and $D \in \mathbb{M}_{m\times n}(\mathbb{R})$ is "diagonal", i.e., the diagonal values of $D_{11}, ..., D_{nn}$ are $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, and they called singular values of A* [1]*.*

Proof: We prove equivalently that there are two unitary matrices $U \in \mathbb{M}_{m\times m}(\mathbb{C})$ and $V \in \mathbb{M}_{n\times n}(\mathbb{C})$ such that

$$U^\star AV = D, \quad \text{where} \quad D_{ii} = \sigma_i \quad \text{are the singular values}$$

We prove it by induction for matrices $A \in \mathbb{M}_{m\times n}(\mathbb{C})$. Specifically, we assume that for every matrix $B \in \mathbb{M}_{(m-1)\times(n-1)}(\mathbb{C})$ there are unitary matrices $U_{m-1} \in \mathbb{M}_{(m-1)\times(m-1)}(\mathbb{C})$ and $V_{n-1} \in \mathbb{M}_{(n-1)\times(n-1)}(\mathbb{C})$ such that $U^\star_{m-1} BV_{n-1} = D_{(m-1)\times(n-1)}$. We want to conclude

[1] If $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ then $A = UDV^\top$, where $U \in \mathbb{M}_{m\times m}(\mathbb{C})$ and $V \in \mathbb{M}_{n\times n}(\mathbb{C})$ are orthogonal

that $U_m^\star A V_n = D$ for unitary matrices U_m, V_m, that are related to U_{m-1}, V_{n-1}. We need to introduce notation: Given a vector $v_1 \in \mathbb{C}^n$ and a matrix $V_2 \in \mathbb{M}_{n\times(n-1)}(\mathbb{C})$ define $(v_1|V_2)$ to be the matrix in $\mathbb{M}_{n\times n}(\mathbb{C})$ constructed by concatenating the column v_1 to the columns V_2. Similarly, given a vector $u_1 \in \mathbb{C}^m$ and a matrix $U_2 \in \mathbb{M}_{m\times(m-1)}(\mathbb{C})$ define $\left(\frac{u_1^\star}{(U_2)^\star}\right)$ to be $(u_1|U_2)^\star$.

Proof sketch: We prove that given a matrix $A \in \mathbb{M}_{m\times n}$ there are two unitary matrices $U = (u_1|U_2) \in \mathbb{M}_{m\times m}$ and $V = (v_1|V_2) \in \mathbb{M}_{n\times n}$ such that

$$(u_1|U_2)^\star \quad A \quad (v_1|V_2) = \left(\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & B \end{array}\right),$$

for some non-negative constant σ_1, and a matrix $B \in \mathbb{M}_{(m-1)\times(n-1)}(\mathbb{C})$. We finish the proof by induction on B.

The proof: Define $\sigma_1 := ||A||_{2\to 2}$, and let v_1 and u_1 be unit vectors (inℓ_2 norm), s.t. $Av_1 = \sigma_1 u_1$. Extend v_1 and u_1 to two unitary matrices: $V = (v_1|V_1)$ and $U = (u_1|U_2)$. Now,

$$U^\star AV = \left(\frac{u_1^\star}{U_2^{m\star}}\right)(A)\,(v_1|V_2^n) = \left(\frac{u_1^\star}{U_2^{m\star}}\right)(Av_1|AV_2^n) = \left(\begin{array}{c|c} u_1^\star Av_1 & u_1^\star AV_2^n \\ \hline U_2^{m\star}Av_1 & U_2^{m\star}AV_2^n \end{array}\right) = \\ = \left(\begin{array}{c|c} \sigma_1 & w \\ \hline 0 & B \end{array}\right).$$

The matrix $U^\star AV$ has the same operator norm as A [2]. We want to show that $w = 0$. Note that

$$\left(\begin{array}{c|c} \sigma_1 & w \\ \hline 0 & B \end{array}\right) \cdot \left(\frac{\sigma_1}{w^\star}\right) = \left(\frac{\sigma_1^2 + ||w||^2}{z}\right).$$

Denote $\phi = (\sigma_1 \,|\, w)^\star$ and $\psi = \left(\sigma_1^2 + ||w||^2 \,|\, z\right)^\star$. Then $||\phi||_2 = \sqrt{\sigma_1^2 + ||w||^2}$ and $||\psi||_2 \geq \sigma_1^2 + ||w||^2$ which together gives $\sigma_1 \geq ||\phi||_2/||\psi||_2 \geq \sqrt{\sigma^2 + ||w||^2} \geq \sigma_1$ and the equality holds only when $w = 0$.

We have shown that

$$U^\star AV = \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & B \end{array}\right).$$

By the induction hypothesis

$$A = (v_1|V_2^n) \cdot \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_{m-1} \end{array}\right) \cdot \left(\begin{array}{c|c} \sigma_1 & \\ \hline & D_{(m-1)\times(n-1)} \end{array}\right) \cdot \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_{m-1} \end{array}\right) \cdot \left(\frac{u_1^\star}{V_2^{n\star}}\right)$$

■

Following the SVD theorem it is straight forward to verify the following statements:

- The Frobenius norm of $A = U^\star DV$ is the square root of the sum of squares of the singular values — $||A||_F = trace(A^\star A) = trace(D^2)$.
- The matrix norm $||A||_{2\to 2}$ is the largest singular value.

[2]For every matrix M and unitary matrices P, Q holds $||PMQ||_{2rightarrow2} = ||M||_{2rightarrow2}$. A proof to be added

1.1 Approximating a matrix A by a matrix of rank k

We can use SVD in the following problem: Give a matrix A and $k \in \mathbb{N}$, find a matrix B with rank $\leq k$, s.t. A and B are as close as possible, i.e. $min\{||A-B||\}$ using some matrix norm. Two natural choices for this norm are $2 \to 2$ and Frobenius norm $||M||_F = \sqrt{\sum M_{i,j}^2}$.

Theorem 2 *Let $A \in \mathbb{M}_{m\times n}(\mathbb{C})$,$m \geq n$ be a matrix. Let $A = UDV^\star$ be its singular value decomposition. Then the optimum of* $\min\{||A-B||\}$ *is achieved with $B = UD^{(k)}V^\star$, where $D^{(k)}_{i,j} = D_{i,j}$ if $i,j \leq k$ and 0 elsewhere.*

Proof: We will prove the theorem for the norm $2 \to 2$. Note that for this choice of B we have $||A-B||_{2\to 2} = \sigma_{k+1}$, and we will show that for any other matrix C of rank $\leq k$ there exist a unit vector z, s.t $||(A-C)z|| \geq \sigma_{k+1}$.

Since $rank\{C\} \leq k$ there is a unit vector z in vector space $span(v_1,\ldots,v_{k+1}) \bigcap Ker(C)$, i.e., $z = \sum_1^{k+1} \alpha_i v_i$, and $Cz = 0$. The matrix A can be written as $A = \sum_{i=1}^m \sigma_i \cdot u_i \otimes \overline{v_i}$ and:

$$(A-C)\cdot z = Az = (\sum_{i=1}^{m} \sigma_i \cdot u_i \otimes \overline{v_i}) \cdot z = \sum_{i=1}^{m} \sigma_i u_i \cdot \langle v_i, z\rangle = \sum_{1}^{k+1} \sigma_i \cdot u_i \cdot \langle v_i, z\rangle\ ,$$

where the last equality follows from the fact the $v_{k+1},...,v_m$ are orthogonal to z. Using the orthogonality of the u_i vector we compute the norm

$$||(A-C)z||_2 = \sum_{i=1}^{k+1} \sigma_i^2 \cdot |\langle v_i, z\rangle|^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} |\langle v_i, z\rangle|^2 = \sigma_{k+1}^2 ||z||^2 = \sigma_{k+1}^2,$$

since the unit vector z is spanned by the orthonormal vectors $v_1,...,v_{k+1}$. The unit vector z is an evidence that lower bounds the matrix norm, i.e., $||A-C||_{2\to 2} \geq \sigma_{k+1}$. ■