CS724: Topics in Algorithms
Singular Values of Matrices

Prof. Dan A. Simovici

A square matrix $A \in \mathbb{C}^{n \times n}$ is *unitarily diagonalizable* if there exists a diagonal matrix $D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{C}^{n \times n}$ and a unitary matrix $X \in \mathbb{C}^{n \times n}$ such that $A = XDX^{\mathsf{H}}$; equivalently, we have $AX = XD$.
Let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the columns of X, then then $A\mathbf{x}_i = d_i\mathbf{x}_i$, which shows that $\mathbf{x}_i$ is a unit eigenvector that corresponds to the eigenvalue d_i for $1 \leqslant i \leqslant n$. Also, we have

$$A = d_1\mathbf{x}_1\mathbf{x}_1^{\mathsf{H}} + \cdots + d_n\mathbf{x}_n\mathbf{x}_n^{\mathsf{H}}.$$

which is the *spectral decomposition of A*. Note that each of the matrices $\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}$ is of rank 1.

The SVD theorem extends this decomposition to rectangular matrices.

Theorem

If $A \in \mathbb{C}^{m\times n}$ is a complex matrix and $\text{rank}(A) = r$, then A can be factored as $A = UDV^H$, where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary matrices,

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & 0 & & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{C}^{m\times n},$$

and $\sigma_1 \geqslant \ldots \geqslant \sigma_r$ are real positive numbers.

Proof

The square matrix $A^{\mathsf{H}}A \in \mathbb{C}^{n\times n}$ is Hermitian, has the same rank as the matrix A and is positive semidefinite. Therefore, there are r positive eigenvalues of this matrix, denoted by $\sigma_1^2, \ldots, \sigma_r^2$, where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be the corresponding pairwise orthogonal, unit eigenvectors in $\mathbb{C}^n$. We have $A^{\mathsf{H}}A\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ for $1 \leqslant i \leqslant r$. Let V be the matrix $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r \ \mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$ obtained by completing the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to an orthogonal basis for $\mathbb{C}^n$. If $V_1 = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r)$ and $V_2 = (\mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$, we can write $V = (V_1 \ V_2)$.
The equalities involving the eigenvectors can now be written as $A^{\mathsf{H}}AV_1 = V_1E^2$, where $E = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.

UMASS BOSTON

Proof cont'd

Define $U_1 = AV_1E^{-1} \in \mathbb{C}^{m\times r}$. We have $U_1^{\mathsf{H}} = E^{-1}V_1^{\mathsf{H}}A^{\mathsf{H}}$, so

$$U_1^{\mathsf{H}}U_1 = E^{-1}V_1^{\mathsf{H}}A^{\mathsf{H}}AV_1E^{-1} = E^{-1}V_1^{\mathsf{H}}V_1E^2E^{-1} = I_r,$$

which shows that the columns of U_1 are pairwise orthogonal unit vectors. Consequently, $U_1^{\mathsf{H}}AV_1E^{-1} = I_r$, so $U_1^{\mathsf{H}}AV_1 = E$.

Proof cont'd

If $U_1 = (\mathbf{u}_1 \ \cdots, \mathbf{u}_r)$, let $U_2 = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m)$ be the matrix whose columns constitute the extension of the set $\{\mathbf{u}_1 \ \cdots, \mathbf{u}_r\}$ to an orthogonal basis of $\mathbb{C}^m$. Define $U \in \mathbb{C}^{m \times m}$ as $U = (U_1 \ U_2)$. Note that

$$\begin{aligned} U^{\mathsf{H}} AV &= \begin{pmatrix} U_1^{\mathsf{H}} \\ U_2^{\mathsf{H}} \end{pmatrix} A(V_1 \ V_2) \\ &= \begin{pmatrix} U_1^{\mathsf{H}} AV_1 & U_1^{\mathsf{H}} AV_2 \\ U_2^{\mathsf{H}} AV_1 & U_2^{\mathsf{H}} AV_2 \end{pmatrix} = \begin{pmatrix} U_1^{\mathsf{H}} AV_1 & U_1^{\mathsf{H}} AV_2 \\ U_2^{\mathsf{H}} AV_1 & U_2^{\mathsf{H}} AV_2 \end{pmatrix} \\ &= \begin{pmatrix} U_1^{\mathsf{H}} AV_1 & O \\ O & O \end{pmatrix} = \begin{pmatrix} E & O \\ O & O \end{pmatrix}, \end{aligned}$$

which is the desired decomposition.

Observe that in the SVD described above (known as the *full SVD*) of A, the diagonal matrix D has the same format as A, while both U and V are square unitary matrices.

Definition

A number $\sigma \in \mathbb{R}_{>0}$ is a *singular value* of a matrix $A \in \mathbb{C}^{m \times n}$ if there exists a pair of vectors $(\mathbf{u}, \mathbf{v}) \in \mathbb{C}^n \times \mathbb{C}^m$ such that

$$A\mathbf{v} = \sigma \mathbf{u} \text{ and } A^{\mathsf{H}}\mathbf{u} = \sigma \mathbf{v}. \tag{1}$$

The vector $\mathbf{u}$ is the *left singular vector* and $\mathbf{v}$ is the *right singular vector* associated to the singular value σ.

If $(\mathbf{u}, \mathbf{v})$ is a pair of vectors associated to σ, then $(a\mathbf{u}, a\mathbf{v})$ is also a pair of vectors associated with σ for every $a \in \mathbb{C}$.
Let $A \in \mathbb{C}^{m\times n}$ and let $A = UDV^{\mathsf{H}}$, where $U \in \mathbb{C}^{m\times m}$, $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{C}^{m\times n}$ and $V \in \mathbb{C}^{n\times n}$. Further, suppose that $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m)$ and $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_n)$.
Since U and V are unitary matrices, we have $U^{\mathsf{H}}\mathbf{u}_j = \mathbf{e}_j$ for $1 \leqslant j \leqslant m$ and $V^{\mathsf{H}}\mathbf{v}_i = \mathbf{e}_i$ for $1 \leqslant i \leqslant n$. Furthermore, $D\mathbf{e}_i = \sigma_i\mathbf{e}_i$ and $D\mathbf{e}_j = \sigma_j\mathbf{e}_j$, which allows us to write:

$$
\begin{aligned}
A\mathbf{v}_i &= UDV^{\mathsf{H}}\mathbf{v}_i = UD\mathbf{e}_i = \sigma_i\mathbf{u}_i, \text{ and} \\
A^{\mathsf{H}}\mathbf{u}_j &= VDU^{\mathsf{H}}\mathbf{u}_j = VD\mathbf{e}_j = \sigma_j\mathbf{v}_j.
\end{aligned}
$$

Thus, the j^{th} column of the matrix U, $\mathbf{u}_j$ and the j^{th} column of the matrix V, $\mathbf{v}_j$ are left and right singular vectors, respectively, associated to the singular value σ_j.

UMASS BOSTON

Corollary

Let $A \in \mathbb{C}^{m \times n}$ *be a matrix and let* $A = UDV^H$ *be the singular value decomposition of* A*. If* $\| \cdot \|$ *is a unitarily invariant norm, then*

$$\| A \| = \| D \| = \| diag(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \| .$$

Proof.

This statement is a direct consequence of the previous Theorem because the matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary. □

Thus, the value of a unitarily invariant norm of a matrix depends only on its singular values. Since $||| \cdot |||_2$ and $\| \cdot \|_F$ are unitarily invariant, the Frobenius norm can be written as

$$\| A \|_F = \sqrt{\sum_{i=1}^{r} \sigma_r^2}.$$

The next definition extends to notion of unitarily equivalent to rectangular matrices.

Definition

Two matrices $A, B \in \mathbb{C}^{m \times n}$ are *unitarily equivalent* (denoted by $A \equiv_u B$) if there exist two unitary matrices W_1 and W_2 such that $A = W_1^{\mathsf{H}} B W_2$.

Theorem

Let A and B be two matrices in $\mathbb{C}^{m\times n}$. If A and B are unitarily equivalent, then they have the same singular values.

Proof.

Suppose that $A \equiv_u B$, that is, $A = W_1^{\mathsf{H}} B W_2$ for some unitary matrices W_1 and W_2. If A has the SVD $A = U^{\mathsf{H}} \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) V$, then

$$B = W_1 A W_2^{\mathsf{H}} = (W_1 U^{\mathsf{H}}) \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)(V W_2^{\mathsf{H}}).$$

Since $W_1 U^{\mathsf{H}}$ and $V W_2^{\mathsf{H}}$ are both unitary matrices, it follows that the singular values of B. □

UMASS BOSTON

Let $\mathbf{v} \in \mathbb{C}^n$ be an eigenvector of the matrix $A^{\mathsf{H}}A$ that corresponds to a non-zero, positive eigenvalue σ^2, that is, $A^{\mathsf{H}}A\mathbf{v} = \sigma^2\mathbf{v}$.
Define $\mathbf{u} = \frac{1}{\sigma}A\mathbf{v}$. We have $A\mathbf{v} = \sigma\mathbf{u}$. Also,

$$A^{\mathsf{H}}\mathbf{u} = A^{\mathsf{H}}\left(\frac{1}{\sigma}A\mathbf{v}\right) = \sigma\mathbf{v}.$$

This implies $AA^{\mathsf{H}}\mathbf{u} = \sigma^2\mathbf{u}$, so $\mathbf{u}$ is an eigenvector of AA^{H} that corresponds to the same eigenvalue σ^2.

Conversely, if $\mathbf{u} \in \mathbb{C}^m$ is an eigenvector of the matrix AA^{H} that corresponds to a non-zero, positive eigenvalue σ^2, we have $AA^{\mathsf{H}}\mathbf{u} = \sigma^2\mathbf{u}$. Thus, if $\mathbf{v} = \frac{1}{\sigma}A\mathbf{u}$ we have $A\mathbf{v} = \sigma\mathbf{u}$ and $\mathbf{v}$ is an eigenvector of $A^{\mathsf{H}}A$ for the eigenvalue σ^2.

The Courant-Fisher Theorem allows the formulation of a similar result for singular values.

Theorem

Let A be a matrix, $A \in \mathbb{C}^{m \times n}$. If $\sigma_1 \geq \sigma_2 \geqslant \cdots \geqslant \sigma_k \geqslant \cdots$ is the non-increasing sequence of singular values of A, then

$$\begin{aligned} \sigma_k &= \min_{\dim(S)=n-k+1} \max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in S \text{ and } \| \mathbf{x} \|_2 = 1\} \\ \sigma_k &= \max_{\dim(T)=k} \min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in T \text{ and } \| \mathbf{x} \|_2 = 1\}, \end{aligned}$$

where S and T range over subspaces of $\mathbb{C}^n$.

Proof

We give the argument only for the second equality of the theorem; the first can be shown in a similar manner.
We saw that σ_k equals the k^{th} largest absolute value of the eigenvalue $|\lambda_k|$ of the matrix $A^{\mathsf{H}}A$. By Courant-Fisher Theorem, we have

$$\begin{aligned}\lambda_k &= \max_{dim(T)=k} \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} A^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in T \text{ and } \parallel \mathbf{x} \parallel_2 = 1\} \\ &= \max_{dim(T)=k} \min_{\mathbf{x}} \{\parallel A\mathbf{x} \parallel_2^2 \mid \mathbf{x} \in T \text{ and } \parallel \mathbf{x} \parallel_2 = 1\},\end{aligned}$$

which implies the second equality of the theorem.

UMASS BOSTON

The theorem can be restated as follows;

Theorem

Let A be a matrix, $A \in \mathbb{C}^{m \times n}$. If $\sigma_1 \geq \sigma_2 \geqslant \cdots \geqslant \sigma_k \geqslant \cdots$ is the non-increasing sequence of singular values of A, then

$$\begin{aligned} \sigma_k &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{k-1} \text{ and } \| \mathbf{x} \|_2 = 1\} \\ &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}} \min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{n-k} \text{ and } \| \mathbf{x} \|_2 = 1\}. \end{aligned}$$

Corollary

The smallest singular value of a matrix $A \in \mathbb{C}^{m \times n}$ *equals*

$$\min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in \mathbb{C}^n \textit{ and } \| \mathbf{x} \|_2 = 1\}.$$

The largest singular value of a matrix $A \in \mathbb{C}^{m \times n}$ *equals*

$$\max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in \mathbb{C}^n \textit{ and } \| \mathbf{x} \|_2 = 1\}.$$

If $A \in \mathbb{C}^{n\times n}$ is an invertible matrix and σ is a singular value of A, then $\frac{1}{\sigma}$ is a singular value of the matrix A^{-1}.

Example

Let

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

be a non-zero vector is $\mathbb{C}^n$, which can also be regarded as a matrix in $\mathbb{C}^{n\times 1}$. The square of a singular value of A is an eigenvalue of the matrix

$$A^{\mathsf{H}}A = \begin{pmatrix} \bar{a}_1 a_1 & \cdots & \bar{a}_n a_1 \\ \bar{a}_1 a_2 & \cdots & \bar{a}_n a_2 \\ \vdots & \ldots & \vdots \\ \bar{a}_1 a_n & \cdots & \bar{a}_n a_n \end{pmatrix}$$

and we have seen that the unique non-zero eigenvalue of this matrix is $\| a \|_2^2$. Thus, the unique singular value of $\mathbf{a}$ is $\| a \|_2$.

Example

Let $A \in \mathbb{R}^{3\times 2}$ be the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrices AA^{H} and $A^{\mathsf{H}}A$ are given by:

$$AA^{\mathsf{H}} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$A^{\mathsf{H}}A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The eigenvalues of $A^{\mathsf{H}}A$ are the roots of the polynomial $\lambda^2 - 4\lambda + 3$, and therefore, they are $\lambda_1 = 3$ and $\lambda_1 = 1$. The eigenvalues of AA^{H} are $3, 1$ and 0.

Unit eigenvectors of $A^{\mathsf{H}}A$ that correspond to 3 and 1 are

$$\begin{aligned} \mathbf{v}_1 &= \alpha_1 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \\ \mathbf{v}_2 &= \alpha_2 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, \end{aligned}$$

respectively, where $\alpha_i \in \{-1, 1\}$ for $i = 1, 2$.

Unit eigenvectors of $A^{\mathsf{H}}A$ that correspond to $3, 1$ and 0 are:

$$\mathbf{u}_1 = \beta_1 \begin{pmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \end{pmatrix}, \mathbf{u}_2 = \beta_2 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, \mathbf{u}_3 = \beta_3 \begin{pmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix},$$

respectively, where $\beta_i \in \{-1, 1\}$ for $i = 1, 2, 3$.

The choice of the columns of the matrices U and V must be done such that for a pair of eigenvectors (u, v) that correspond to a singular values σ we have $\mathbf{v} = \frac{1}{\sigma} A^{\mathsf{H}} \mathbf{u}$ or, equivalently, $\mathbf{u} = \frac{1}{\sigma} A\mathbf{v}$. For instance, if we choose $\alpha_1 = \alpha_2 = 1$, then

$$\mathbf{v}_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix},$$

and $\mathbf{u}_1 = \frac{1}{\sqrt{3}} A\mathbf{v}_1$ and $\mathbf{u}_2 = A\mathbf{v}_2$, that is

$$\mathbf{u}_1 = \begin{pmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix},$$

which means that $\beta_1 = 1$ and $\beta_2 = -1$; the value of β_3 that corresponds to the the eigenvalue of 0 can be chosen arbitrarily.

UMASS BOSTON

Thus, an SVD of A is

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

- The singular values of a matrix $A \in \mathbb{C}^{m \times n}$ are uniquely determined.
- The matrices U and V of the SVD of A are not unique. Once we choose a column of the matrix V for a singular value σ, the corresponding column of U is determined by $\mathbf{u} = \frac{1}{\sigma} A\mathbf{v}$.

A variant of the SVD Decomposition Theorem is given next.

Corollary

(The Thin SVD Decomposition Corollary) *Let* $A \in \mathbb{C}^{m \times n}$ *be a matrix having non-zero singular values* $\sigma_1, \sigma_2, \ldots, \sigma_r$*, where* $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ *and* $r \leqslant \min\{m, n\}$*. Then, A can be factored as* $A = UDV^H$*, where* $U \in \mathbb{C}^{m \times r}$ *and* $V \in \mathbb{C}^{n \times r}$ *are matrices having orthonormal sets of columns and D is the diagonal matrix*

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

The decomposition described in above is known as a *thin SVD decomposition* of the matrix A.

Example

The thin SVD decomposition of the matrix A,

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

is

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Since U and V in the thin SVD have orthonormal columns it is easy to see that

$$U^{\mathsf{H}} U = V^{\mathsf{H}} V = I_p.$$

Lemma

Let $D \in \mathbb{R}^{n\times n}$ *be a diagonal matrix, where* $D = diag(\sigma_1, \ldots, \sigma_r)$ *and* $\sigma_1 \geqslant \cdots \geq \sigma_r$. *Then, we have* $\|D\|_2 = \sigma_1$, *and* $\| D \|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$.

Proof

By the definition of $\|D\|_2$ we have:

$$\begin{aligned} \|D\|_2 &= \max\{\| D\mathbf{x} \|_2 \mid \| \mathbf{x} \| = 1\} \\ &= \max\{\sqrt{\sum_{i=1}^{r} \sigma_i^2 |x_i|^2} \mid \sum_{i=1}^{n} |x_i|^2 = 1\}. \end{aligned}$$

Since

$$\sum_{i=1}^{r} \sigma_i^2 |x_i|^2 \leq \sigma_1^2 \left(\sum_{i=1}^{r} |x_i|^2 \right) \leqslant \sigma_1^2,$$

because $\sum_{i=1}^{n} |x_i|^2 = 1$, it follows that

$$\max\left\{ \sqrt{\sum_{i=1}^{r} \sigma_i^2 |x_i|^2} \,\middle|\, \sum_{i=1}^{n} |x_i|^2 = 1 \right\} = \sigma_1.$$

The second part is immediate.

Theorem

Let $A \in \mathbb{C}^{m\times n}$ be a matrix whose singular values are $\sigma_1 \geqslant \cdots \geqslant \sigma_r$. Then $\|A\|_2 = \sigma_1$, and $\| A \|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$.

Proof.

Suppose that the SVD of A is $A = UDV^{\mathsf{H}}$, where U and V are unitary matrices. Then, we have:

$$\begin{aligned} \|A\|_2 &= \|UDV^{\mathsf{H}}\|_2 = \|D\|_2 = \sigma_1, \\ \| A \|_F &= \| UDV^{\mathsf{H}} \|_F = \| D \|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}. \end{aligned}$$

□

UMASS BOSTON

Corollary

If $A \in \mathbb{C}^{m \times n}$ is a matrix, then $\|A\|_2 \leqslant \| A \|_F \leq \sqrt{n}\|A\|_2$.

Proof.

Suppose that $\sigma_1(A)$ is the largest of the singular values of A. Then, since $\| A \|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$, we have

$$\sigma_1(A) \leqslant \| A \|_F \leqslant \sqrt{n \max_i \sigma_j(A)^2} = \sigma_1(A)\sqrt{n},$$

which is desired double inequality. □

Let $A = UDV^{\mathsf{H}}$ be an SVD of A. If we write U and V using their columns as

$$U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m), V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_n),$$

then A can be written as

$$\begin{aligned} A &= UDV^{\mathsf{H}} \\ &= (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n) \begin{pmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^{\mathsf{H}} \\ \vdots \\ \mathbf{v}_m^{\mathsf{H}} \end{pmatrix} \\ &= (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m) \begin{pmatrix} \sigma_1 \mathbf{v}_1^{\mathsf{H}} \\ \vdots \\ \sigma_r \mathbf{v}_p^{\mathsf{H}} \end{pmatrix} \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_p^{\mathsf{H}}. \end{aligned} \quad (2)$$

UMASS BOSTON

Since $\mathbf{u}_i \in \mathbb{C}^m$ and $\mathbf{v}_i \in \mathbb{C}^n$, each of the matrices $\mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}$ is a $m \times n$ matrix of rank 1. Thus, the SVD yields an expression of A as a sum of r matrices of rank 1, where r is the number of non-zero singular values of A.

Theorem

The rank-1 matrices of the form $\mathbf{u}_i\mathbf{v}_i^H$, *where* $1 \leq i \leqslant r$ *are pairwise orthogonal. Moreover,* $\| \mathbf{u}_i\mathbf{v}_i^H \|_F = 1$ *for* $1 \leqslant i \leqslant r$.

Proof

For $i \neq j$ and $1 \leqslant i, j \leqslant r$ we have:

$$trace\left(\mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} (\mathbf{u}_j \mathbf{v}_j^{\mathsf{H}})^{\mathsf{H}}\right) = trace\left(\mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \mathbf{v}_j \mathbf{u}_j\right) = 0,$$

because the vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal. Thus, $(\mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}, \mathbf{u}_j \mathbf{v}_j^{\mathsf{H}}) = 0$. Therefore, we have

$$\begin{aligned} \| \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \|_F^2 &= trace((\mathbf{u}_i \mathbf{v}_i^{\mathsf{H}})^{\mathsf{H}} \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}) \\ &= trace(\mathbf{v}_i \mathbf{u}_i^{\mathsf{H}} \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}) = 1, \end{aligned}$$

because the matrices U and V are unitary.

Theorem

Let $A \in \mathbb{C}^{m \times n}$ be a matrix that has the singular value decomposition $A = UDV^H$. If $rank(A) = r$, then the first r columns of U form an orthonormal basis for $Ran(A)$, and the last $n - r$ columns of V constitute an orthonormal basis for $NullSp(A)$.

Proof

Since both U and V are unitary matrices, it is clear that $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, the set of the first r columns of U, and $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$, the set of the last $n - r$ columns of V, are linearly independent sets. Thus, we only need to show that $\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \text{Ran}(A)$ and $\langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle = \text{NullSp}(A)$.
We have

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^{\mathsf{H}}.$$

If $\mathbf{t} \in \text{Ran}(A)$, then $\mathbf{t} = A\mathbf{s}$ for some $\mathbf{s} \in \mathbb{C}^n$. Therefore, $\mathbf{t} = \sigma_1 \mathbf{u}_1 (\mathbf{v}_1^{\mathsf{H}} \mathbf{s}) + \cdots + \sigma_r \mathbf{u}_r (\mathbf{v}_r^{\mathsf{H}} \mathbf{s})$, and, since the every product $\mathbf{v}_j^{\mathsf{H}} \mathbf{s}$ is a scalar for $1 \leqslant j \leqslant r$, it follows that $\mathbf{t} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$, so $\text{Ran}(A) \subseteq \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$.

Proof cont'd

To prove the reverse inclusion note that

$$A\left(\frac{1}{\sigma_i}\mathbf{v}_i\right) = \mathbf{u}_i,$$

for $1 \leqslant i \leqslant r$, due to the orthogonality of the columns of V. Thus, $\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \mathrm{Ran}(A)$.
Note that

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_p^{\mathsf{H}}$$

implies that $A\mathbf{v}_j = 0$ for $r+1 \leqslant j \leqslant n$, so $\langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle \subseteq \mathrm{NullSp}(A)$. Conversely, suppose that $A\mathbf{r} = \mathbf{0}$. Since the columns of V form a basis of $\mathbb{C}^n$ we have $\mathbf{r} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, so $A\mathbf{r} = a_1 A\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r = \mathbf{0}$. The linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ implies $a_1 = \cdots = a_r = 0$, so $\mathbf{r} = a_{r+1}\mathbf{v}_{r+1} + \cdots + a_n\mathbf{v}_n$, which shows that $\mathrm{NullSp}(A) \subseteq \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle$. Thus, $\mathrm{NullSp}(A) = \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle$.

Corollary

Let $A \in \mathbb{C}^{m\times n}$ be a matrix that has the singular value decomposition $A = UDV^H$. If $\text{rank}(A) = r$, then the first r transposed columns of V form an orthonormal basis for the subspace of $\mathbb{R}^n$ generated by the rows of A.

Proof.

This statement follows immediately from the theorem applied to A^{H}. □

The SVD allows us to find the best approximation of of a matrix by a matrices of limited rank.

Lemma

Let $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^H$ *be the SVD of a matrix* $A \in \mathbb{R}^{m \times n}$*, where* $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$*. For every* k*,* $1 \leqslant k \leqslant r$ *the matrix* $B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$ *has rank* k*.*

Proof

The null space of the matrix $B(k)$ consists of those vectors $\mathbf{x}$ such that $\sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \mathbf{x} = \mathbf{0}$. The linear independence of the vectors $\mathbf{u}_i$ and the fact that $\sigma_i > 0$ for $1 \leqslant i \leqslant r$ implies the equalities $\mathbf{v}_i^{\mathsf{H}} \mathbf{x} = \mathbf{0}$ for $1 \leqslant i \leqslant r$. Thus,

$$\mathrm{NullSp}(B(k)) = \mathrm{NullSp}\left(\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{pmatrix}\right).$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent it follows that $\dim(\mathrm{NullSp}(B(k)) = n - k$, which implies $rank(B(k)) = k$ for $1 \leqslant k \leqslant r$.

Theorem

(Eckhart-Young Theorem) *Let* $A \in \mathbb{C}^{m \times n}$ *be a matrix whose sequence of non-zero singular values is* $(\sigma_1, \ldots, \sigma_r)$. *Assume that* $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$ *and that* A *can be written as*

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^H.$$

Let $B(k) \in \mathbb{C}^{m \times n}$ *be the matrix defined by*

$$B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^H.$$

If $r_k = \inf\{\|A - X\|_2 \mid X \in \mathbb{C}^{m \times n} \text{ and } \operatorname{rank}(X) \leqslant k\}$, *then*

$$\|A - B(k)\|_2 = r_k = \sigma_{k+1},$$

for $1 \leqslant k \leqslant r$, *where* $\sigma_{r+1} = 0$ *and* $B(k)$ *is the best approximation of* A *among the matrices of rank no larger than* k *in the sense of the norm* $\|\cdot\|_2$.

Proof

Observe that

$$A - B(k) = \sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}},$$

and the largest singular value of the matrix $\sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}$ is σ_{k+1}.
Therefore,

$$\|A - B(k)\|_2 = \sigma_{k+1}.$$

for $1 \leqslant k \leqslant r$.
We prove now that for every matrix $X \in \mathbb{C}^{m \times n}$ such that $rank(X) \leqslant k$, we have $\|A - X\|_2 \geqslant \sigma_{k+1}$. Since $\dim(\text{NullSp}(X)) = n - rank(X)$, it follows that $\dim(\text{NullSp}(X)) \geqslant n - k$. If T is the subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$, we have $\dim(T) = k + 1$. Since $\dim(\text{NullSp}(X)) + \dim(T) > n$, the intersection of these subspaces contains a non-zero vector and, without loss of generality, we can assume that this vector is a unit vector $\mathbf{x}$.

Proof cont'd

We have $\mathbf{x} = a_1\mathbf{v}_1 + \cdots a_k\mathbf{v}_k + a_{k+1}\mathbf{v}_{k+1}$ because $\mathbf{x} \in T$. The orthogonality of $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$ implies $\| \mathbf{x} \|_2^2 = \sum_{i=1}^{k+1} |a_i|^2 = 1$. Since $\mathbf{x} \in \text{NullSp}(X)$, we have $X\mathbf{x} = \mathbf{0}$, so

$$(A - X)\mathbf{x} = A\mathbf{x} = \sum_{i=1}^{k+1} a_i A\mathbf{v}_i = \sum_{i=1}^{k+1} a_i \sigma_i \mathbf{u}_i.$$

Thus, we have

$$\|(A - X)\mathbf{x}\|_2^2 = \sum_{i=1}^{k+1} |\sigma_i a_i|^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} |a_i|^2 = \sigma_{k+1}^2,$$

because $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are also orthonormal. This implies $\|A - X\|_2 \geqslant \sigma_{k+1} = \|A - B(k)\|_2$.

Singular vector decompositions of matrices can be computed using the function `svd`. Its standard usage for an $n \times p$-matrix `x` is

```
svd(x, nu, nv)
```

where `nu` is the number of left singular vectors to be computed (which must be between 0 and `n`) and `nv` is the number of right singular vectors to be computed (between 0 and `p`). The arguments `nu` and `nv` are optional and have the default values n and p, respectively.

```
> svd(x)
$d
[1] 9.5255181 0.5143006

$u
           [,1]       [,2]
[1,] -0.6196295 -0.7848945
[2,] -0.7848945  0.6196295

$v
           [,1]       [,2]
[1,] -0.2298477  0.8834610
[2,] -0.5247448  0.2407825
[3,] -0.8196419 -0.4018960
```