CS724: Topics in Algorithms
Singular Values
Slide Set 8

Prof. Dan A. Simovici

The singular value decomposition has been described as the “Swiss Army knife of matrix decompositions” due to its many applications in the study of matrices; from our point of view, singular value decomposition is relevant for dimensionality reduction techniques in data mining.

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

$$
\begin{aligned}
A &= (\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_n)\mathrm{diag}(d_1, \ldots, d_n) \begin{pmatrix} \boldsymbol{x}_1^{\mathsf{H}} \\ \vdots \\ \boldsymbol{x}_n^{\mathsf{H}} \end{pmatrix} \\
&= d_1\boldsymbol{x}_1\boldsymbol{x}_1^{\mathsf{H}} + \cdots + d_n\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}}.
\end{aligned}
$$

This is the *spectral decomposition of A*. Note that $\mathrm{rank}(\boldsymbol{x}_i\boldsymbol{x}_i^{\mathsf{H}}) = 1$ for $1 \leqslant i \leqslant n$.

UMASS BOSTON

The SVD theorem extends this decomposition to rectangular matrices.

Theorem

(SVD Theorem) *If $A \in \mathbb{C}^{m\times n}$ is a 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 $A^{\mathsf{H}}A \in \mathbb{C}^{n\times n}$ 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 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ be the corresponding pairwise orthogonal, unit eigenvectors in $\mathbb{C}^n$ and let

$$V_1 = (\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_r),$$

$V = (\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_r \ \boldsymbol{v}_{r+1} \ \cdots \ \boldsymbol{v}_n)$ be the matrix obtained by completing the set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$ to an orthogonal basis for $\mathbb{C}^n$, and let $V_2 = (\boldsymbol{v}_{r+1} \ \cdots \ \boldsymbol{v}_n)$.
We can write $V = (V_1 \ V_2)$.
Note that we have $A^{\mathsf{H}}A\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i$ for $1 \leqslant i \leqslant r$.

Proof cont'd

The equalities $A^{\mathsf{H}}A\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i$ for $1 \leqslant i \leqslant r$ involving the eigenvectors can now be written as

$$A^{\mathsf{H}}AV_1 = V_1E^2,$$

where $E = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.
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 $U_1 = (\boldsymbol{u}_1 \ \cdots, \boldsymbol{u}_r)$, let $U_2 = (\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m)$ be the matrix whose columns constitute the extension of the set $\{\boldsymbol{u}_1 \ \cdots, \boldsymbol{u}_r\}$ 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

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

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

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

Note that if $(\boldsymbol{u}, \boldsymbol{v})$ is a pair of vectors associated to σ, then $(a\boldsymbol{u}, a\boldsymbol{v})$ 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 $V \in \mathbb{C}^{n \times n}$.
Note that

$$
\begin{aligned}
A\boldsymbol{v}_j &= UDV^{\mathsf{H}}\boldsymbol{v}_j = UD\boldsymbol{e}_j \\
&\quad \text{(because } V \text{ is a unitary matrix)} \\
&= \sigma_j U\boldsymbol{e}_j = \sigma_j \boldsymbol{u}_j
\end{aligned}
$$

and

$$
\begin{aligned}
A^{\mathsf{H}}\boldsymbol{u}_j &= VD^{\mathsf{H}}U^{\mathsf{H}}\boldsymbol{u}_j = VDU^{\mathsf{H}}\boldsymbol{u}_j \\
&\quad VD\boldsymbol{e}_j \\
&\quad \text{(because } U \text{ is a unitary matrix)} \\
&= \sigma_j V\boldsymbol{e}_j = \sigma_j \boldsymbol{v}_j.
\end{aligned}
$$

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

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 follows from the fact that the matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary. □

In other words, the value of a unitarily invariant norm of a matrix depends only on its singular values. As we saw, $||| \cdot |||_2$ and $\| \cdot \|_F$ are unitarily invariant. Therefore, the Frobenius norm can be written as

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

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 $A = W_1^{\mathsf{H}} B W_2$.
Clearly, if $A \sim_u B$, then $A \equiv_u B$.

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 are the same as the singular values of A.

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

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

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

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

Theorem

Let A be a matrix, $A \in \mathbb{C}^{m \times n}$. If $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_k \geqslant \cdots$ 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

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

which implies the second equality of the theorem.

Theorem

Let A be a matrix, $A \in \mathbb{C}^{m \times n}$. If $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_k \geqslant \cdots$ 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

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

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

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

The SVD theorem can also be proven by induction on $q = \min\{m, n\}$. In the base case, $q = 1$, we have $A \in \mathbb{C}^{1\times 1}$, or $A \in \mathbb{C}^{m\times 1}$, or $A \in \mathbb{C}^{1\times n}$. Suppose, for example, that $A = \boldsymbol{a} \in \mathbb{C}^{m\times 1}$, where

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

and let $a = \| \boldsymbol{a} \|_2$. We seek $U \in \mathbb{C}^{m\times m}$, $V = (v) \in \mathbb{C}^{1\times 1}$ such that

$$\boldsymbol{a} = UDv,$$

where $D \in \mathbb{C}^{m\times 1}$ is

$$D = \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{C}^{m\times 1}.$$

The role of the matrix U is played by any unitary matrix which has the first column equal to

$$\begin{pmatrix} \frac{a_1}{a} \\ \frac{a_2}{a} \\ \vdots \\ \frac{a_n}{a} \end{pmatrix},$$

and we can adopt $v = 1$. The remaining base subcases can be treated in a similar manner.

Suppose now that the statement holds when at least one of the numbers m and n is less than q and let us prove the assertion when at least one of m and n is less than $q+1$.
Let $\boldsymbol{u}_1$ be a unit eigenvector of AA^{H} that corresponds to the eigenvalue σ_1^2. We have:

$$AA^{\mathsf{H}}\boldsymbol{u}_1 = \sigma_1^2\boldsymbol{u}.$$

Define $\boldsymbol{v}_1 = \frac{1}{\sigma_1}A^{\mathsf{H}}\boldsymbol{u}_1$. Note that:

$$\begin{aligned} \parallel \boldsymbol{v}_1 \parallel^2 &= \boldsymbol{v}_1^{\mathsf{H}}\boldsymbol{v}_1 = \frac{1}{\sigma_1^2}\boldsymbol{u}_1^{\mathsf{H}}AA^{\mathsf{H}}\boldsymbol{u}_1 \\ &= \boldsymbol{u}_1^{\mathsf{H}}\boldsymbol{u}_1 = \parallel \boldsymbol{u}_1 \parallel^2 = 1. \end{aligned}$$

Also, we have:

$$A\boldsymbol{v}_1 = \frac{1}{\sigma_1} AA^{\mathsf{H}} \boldsymbol{u}_1 = \sigma_1 \boldsymbol{u}_1,$$

which shows that $(\boldsymbol{v}_1, \boldsymbol{u}_1)$ is a pair of singular vectors corresponding to the singular value σ_1.
We have also

$$\boldsymbol{u}_1^{\mathsf{H}} A^{\mathsf{H}} \boldsymbol{v}_1 = \frac{1}{\sigma_1} \boldsymbol{u}_1^{\mathsf{H}} AA^{\mathsf{H}} \boldsymbol{u}_1 = \sigma_1.$$

UMASS BOSTON

Define $U = (\boldsymbol{u}_1 \ U_1)$ and $V = (\boldsymbol{v}_1 \ V_1)$ as unitary matrices having $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ as their first columns, respectively. Then,

$$
\begin{aligned}
U^{\mathsf{H}} A^{\mathsf{H}} V &= \begin{pmatrix} \boldsymbol{u}_1^{\mathsf{H}} \\ U_1^{\mathsf{H}} \end{pmatrix} A^{\mathsf{H}} \begin{pmatrix} \boldsymbol{v}_1 & V_1 \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{u}_1^{\mathsf{H}} A^{\mathsf{H}} \\ U_1^{\mathsf{H}} A^{\mathsf{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 & V_1 \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{u}_1^{\mathsf{H}} A^{\mathsf{H}} \boldsymbol{v}_1 & \boldsymbol{u}_1^{\mathsf{H}} A V_1 \\ U_1^{\mathsf{H}} A \boldsymbol{v}_1 & U_1^{\mathsf{H}} A V_1 \end{pmatrix}
\end{aligned}
$$

Since U is a unitary matrix every column of U_1 is orthogonal to $\boldsymbol{u}_1$. Therefore,

$$U_1^{\mathsf{H}} A \boldsymbol{v}_1 = \frac{1}{\sigma_1} U_1^{\mathsf{H}} A A^{\mathsf{H}} \boldsymbol{u}_1 = \sigma_1 U_1^{\mathsf{H}} \boldsymbol{u}_1 = \mathbf{0},$$

and, similarly,

$$\boldsymbol{u}_1^{\mathsf{H}} A^{\mathsf{H}} V_1 = \sigma_1 \boldsymbol{v}_1^{\mathsf{H}} V_1 = \mathbf{0}',$$

because $\boldsymbol{v}_1$ is orthogonal on all columns of V_1. Thus,

$$U^{\mathsf{H}} A V = \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & U_1^{\mathsf{H}} A V_1 \end{pmatrix}.$$

The matrix $U_1^{\mathsf{H}} A V_1$ has fewer rows and columns than $U^{\mathsf{H}} A V$, so we can apply the inductive hypothesis to $B = U_1^{\mathsf{H}} A V_1$. Therefore, by the inductive hypothesis, B can be written as $B = X D Y^{\mathsf{H}}$, where X and Y are unitary matrices and D is a diagonal matrix.

This allows us to write

$$U^{\mathsf{H}}AV = \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & XDY^{\mathsf{H}} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & X \end{pmatrix} \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & D \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Y^{\mathsf{H}} \end{pmatrix}.$$

Since the matrices

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & X \end{pmatrix} \text{ and } \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Y^{\mathsf{H}} \end{pmatrix}$$

are unitary, we obtain the desired conclusion.

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}.

Reminder

Let

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{C}^n - \{\mathbf{0}\}.$$

Recall that the matrix $\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}$ has the same non-zero eigenvalues as the matrix $\boldsymbol{a}^{\mathsf{H}}\boldsymbol{a}$. Since

$$\boldsymbol{a}^{\mathsf{H}}\boldsymbol{a} = \bar{a}_1 a_1 + \cdots + \bar{a}_n a_n \in \mathbb{C}$$

is a scalar, its unique eigenvalue is $|a_1|^2 + \cdots + |a_n|^2 = \parallel \boldsymbol{a} \parallel^2$, hence the matrix $\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}$ has a unique eigenvalue is $\parallel \boldsymbol{a} \parallel^2$.

Example

Let

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

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

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

and we have seen that the unique non-zero eigenvalue of this matrix is $\parallel a \parallel_2^2$. Thus, the unique singular value of $\boldsymbol{a}$ is $\parallel a \parallel_2$.

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:

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

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 $\lambda_2 = 1$. The eigenvalues of AA^{H} are $3, 1$ and 0.

Example

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

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

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

Example

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

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

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

Example

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 $\boldsymbol{v} = \frac{1}{\sigma} A^{\mathsf{H}} \boldsymbol{u}$ or, equivalently, $\boldsymbol{u} = \frac{1}{\sigma} A \boldsymbol{v}$. For instance, if we choose $\alpha_1 = \alpha_2 = 1$, then

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

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

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

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

Example

Thus, an SVD of A is:

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

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*

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

Proof.

The statement is an immediate consequence of the theorem. □

The decomposition described 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

$$U^{\mathsf{H}} U = V^{\mathsf{H}} V = I_p. \tag{2}$$

Lemma

Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix, where $D = diag(\sigma_1, \ldots, \sigma_r)$ and $\sigma_1 \geqslant \cdots \geqslant \sigma_r$. 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:

$$\|D\|_2 = \max\{\| D\mathbf{x} \|_2 \mid \| \mathbf{x} \| = 1\} = \max\left\{ \sqrt{\sum_{i=1}^r \sigma_i^2 |x_i|^2} \,\middle|\, \sum_{i=1}^n |x_i|^2 = 1 \right\}.$$

Since $\sum_{i=1}^n |x_i|^2 = 1$, we have: $\sum_{i=1}^r \sigma_i^2 |x_i|^2 \leqslant \sigma_1^2 \left(\sum_{i=1}^r |x_i|^2\right) \leqslant \sigma_1^2$.
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, by previous results, we have:

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

□

Corollary

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

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. □

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be an invertible matrix. If the singular values of A are $\sigma_1 \geqslant \cdots \geqslant \sigma_n > 0$, then

$$cond(A) = \frac{\sigma_1}{\sigma_n}.$$

Proof.

We have shown that $\|A\|_2 = \sigma_1$. Since the singular values of A^{-1} are

$$\frac{1}{\sigma_n} \geqslant \cdots \geqslant \frac{1}{\sigma_1}$$

it follows that $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$. The desired equality follows immediately. □

Corollary

Let $A \in \mathbb{C}^{n\times n}$ be an invertible matrix. We have $cond(A^H A) = (cond(A))^2$.

Proof.

Let σ be a singular value of A and let $\boldsymbol{u}, \boldsymbol{v}$ be two left and right singular vectors corresponding to σ, respectively. We have

$$A\boldsymbol{v} = \sigma\boldsymbol{u} \text{ and } A^{\mathsf{H}}\boldsymbol{u} = \sigma\boldsymbol{v}.$$

This implies $A^{\mathsf{H}}A\boldsymbol{v} = \sigma A^{\mathsf{H}}\boldsymbol{u} = \sigma^2\boldsymbol{v}$, which shows that the singular values of the matrix $A^{\mathsf{H}}A$ are the squares of the singular values of A, which produces the desired conclusion. □

Let $A = UDV^{\mathsf{H}}$ be an SVD of A. If we write U and V using their columns as $U = (\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_m)$ and $V = (\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_n)$, then A can be written as:

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

UMASS BOSTON

Since $\boldsymbol{u}_i \in \mathbb{C}^m$ and $\boldsymbol{v}_i \in \mathbb{C}^n$, each of the matrices $\boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{H}}$ 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 $1 \leqslant i \leqslant r$ 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:

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

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

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

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 $range(A)$, and the last $n - r$ columns of V constitute an orthonormal basis for $null(A)$.

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 $\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \text{range}(A)$ and $\langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle = \text{null}(A)$.
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 $\mathbf{t} \in \text{range}(A)$, 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 $\text{range}(A) \subseteq \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$.

Proof cont'd

To prove the reverse inclusion note that

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

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

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 previous theorem applied to A^{H}. $\square$

The singular value decomposition of a matrix can be computed in MATLAB using the function svd. To illustrate its use in the simplest form consider the matrix

$$A = \begin{pmatrix} \frac{1}{3} & \frac{4}{3} & \frac{5}{3} \\ \frac{11}{5} & \frac{4}{5} & 3 \\ \frac{1}{7} & \frac{4}{7} & \frac{5}{7} \\ \frac{1}{9} & \frac{4}{9} & \frac{5}{9} \end{pmatrix},$$

defined in MATLAB by

```
A = [1/3 4/3 5/3 ; 11/5 4/5 3 ; 1/7 4/7 5/7 ; 1/9 4/9 5/9]
```

A call the `svd(A)` function yields a vector containing the singular values of A, as in

```
>> svd(A)
ans =
    4.3674
    1.2034
    0.0000
```

It is interesting to note that $\text{rank}(A) = 2$ since the last column of A is the sum of the first two columns. Thus, we would expect to see two non-zero singular values.

To compute $\|A\|_2$, which equals the largest singular value of A we can use `max(svd(A))`.

Another variant of the `svd` function, `[U,S,V] = svd(A)` yields a diagonal matrix S, of the same format as A and with nonnegative diagonal elements in decreasing order, and unitary matrices U and V so that $A = USV^{\mathsf{H}}$. For the matrix A shown above we obtain:

Example

```
>> [U,S,V] = svd(A)
U =
   -0.4487    0.7557    0.4769    0.0161
   -0.8599   -0.5105    0.0000   -0.0000
   -0.1923    0.3239   -0.6726   -0.6370
   -0.1496    0.2519   -0.5658    0.7707
S =
    4.3674         0         0
         0    1.2034         0
         0         0    0.0000
         0         0         0
V =

   -0.4775   -0.6623    0.5774
   -0.3349    0.7447    0.5774
   -0.8123    0.0823   -0.5774
```

The "economical form" of the svd function is

```
[U,S,V] = svd(A,'econ')
```

If $A \in \mathbb{R}^{m\times n}$ and $m > n$, only the first n columns of U are computed and $S \in \mathbb{R}^{n\times n}$. If $m < n$ only the first m columns of V are computed.

Example

Starting from the matrix

$$A = \begin{pmatrix} 18 & 8 & 20 \\ -4 & 20 & 1 \\ 25 & 8 & 27 \\ 9 & 4 & 10 \end{pmatrix} \in \mathbb{R}^{4\times 3}$$

a call the economical variant of the svd function yields

Example

```
>> [U,D,V] = svd(A,'econ')
U =
   -0.5717   -0.0211    0.8095
   -0.0721   -0.9933   -0.0669
   -0.7656    0.1133   -0.4685
   -0.2859   -0.0105   -0.3474

S =
   49.0923         0         0
         0   20.2471         0
         0         0    0.0000


V =
   -0.6461    0.3127    0.6963
   -0.2706   -0.9468    0.1741
   -0.7137    0.0760   -0.6963
```

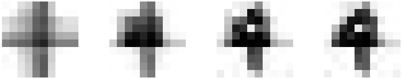
The function `svapprox` given below computes the successive approximations $B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}$ of a matrix $A \in \mathbb{R}^{m \times n}$ having the SVD $A = UDV^{\mathsf{H}}$ and produces a three-dimensional array $C \in \mathbb{R}^{m \times n \times r}$, where r is the numerical rank of A and $C(:,:,k) = B(k)$ for $1 \leqslant k \leqslant r$.

```
function [C] = svapprox(A)
%SVAPPROX computes the successive approximations
%      of A using the singular component decomposition.
%      The number of approximations equals the
%          numerical rank of A.

% determine the format of A and its numerical rank
[m, n] = size(A);
r = rank(A,10^(-5));
% compute the SVD of A
[U,D,V] = svd(A);
C = zeros(m,n,r);
C(:,:,1) = D(1,1) * U(:,1) * (V(:,1))';
for k=2:r
    C(:,:,k) = D(k,k) * U(:,k) * (V(:,k))' + C(:,:,k-1);
end;
```

In the next figure we have an image of the digit 4 created from a `pgm` file that contains the representation of this digit.
The numerical rank of the matrix A introduced in the example mentioned above is 8. Therefore, the array C computed by `C = svapprox(A)` consists of 8 matrices. To represent these matrices in the `pgm` format we cast the components of C to integers of the type `uint8` using `D = min(16,uint8(C))`. Thus, $D(:,:,j)$ contains the rounded j^{th} approximation of A.

The images for the first four approximations are represented next:



(a) (b) (c) (d)

Successive Approximations of A.

Note that the digit four is easily recognizable beginning with the second approximation.