CS724: Topics in Algorithms
Spectral Properties of Matrices - 1

Prof. Dan A. Simovici

University of Massachusetts Boston

February 3, 2014

Let $A \in \mathbb{C}^{n \times n}$ be a square matrix.

- An *eigenpair* of A is a pair $(\lambda, \mathbf{x}) \in \mathbb{C} \times (\mathbb{C}^n - \{\mathbf{0}\})$ such that $A\mathbf{x} = \lambda\mathbf{x}$.
- We refer to λ is an *eigenvalue* and to $\mathbf{x}$ is an *eigenvector*.
- The set of eigenvalues of A is the *spectrum* of A and will be denoted by $\text{spec}(A)$.

If $(\lambda, \mathbf{x})$ is an eigenpair of A, the linear system $A\mathbf{x} = \lambda\mathbf{x}$ has a non-trivial solution in $\mathbf{x}$. An equivalent homogeneous system is $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$ and this system has a non-trivial solution only if $\det(\lambda I_n - A) = 0$.

Definition

The *characteristic polynomial* of the matrix A is the polynomial p_A defined by $p_A(\lambda) = \det(\lambda I_n - A)$ for $\lambda \in \mathbb{C}$.

Thus, the eigenvalues of A are the roots of the characteristic polynomial of A.

Lemma

Let $A = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n) \in \mathbb{C}^n$ *and let* B *be the matrix obtained from* A *by replacing the column* $\mathbf{a}_j$ *by* $\mathbf{e}_j$*. Then, we have*

$$\det(B) = \det\left(A\begin{bmatrix} 1 & \cdots & j-1 & j+1 & \cdots & n \\ 1 & \cdots & j-1 & j+1 & \cdots & n \end{bmatrix}\right).$$

If B is obtained from A by replacing the columns $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_k}$ by $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}$ and $\{i_1, \ldots, i_p\} = \{1, \ldots, n\} - \{j_1, \ldots, j_k\}$, then

$$\det(B) = \det\left(A\begin{bmatrix} i_1 & \cdots & i_p \\ i_1 & \cdots & i_p \end{bmatrix}\right). \quad (1)$$

In other words, $\det(B)$ equals a principal p-minor of A.

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a matrix. Its characteristic polynomial p_A can be written as

$$p_A(\lambda) = \sum_{k=0}^{n} (-1)^k a_k \lambda^{n-k},$$

where a_k is the sum of the principal minors of order k of A.

Proof

$$p_A(\lambda) = \det(\lambda I_n - A) = (-1)^n \det(A - \lambda I_n)$$

can be written as a sum of 2^n determinants of matrices obtained by replacing each subset of the columns of A by the corresponding subset of columns of $-\lambda I_n$.

If the subset of columns of $-\lambda I_n$ involved are $-\lambda \mathbf{e}_{j_1}, \ldots, -\lambda \mathbf{e}_{j_k}$ the result of the substitution is $(-1)^k \lambda^k \det \left(A \begin{bmatrix} i_1 \cdots i_p \\ i_1 \cdots i_p \end{bmatrix} \right)$, where $\{i_1, \ldots, i_p\} = \{1, \ldots, n\} - \{j_1, \ldots, j_k\}$. The total contribution of sets of k columns of $-\lambda I_n$ is $(-1)^k \lambda^k a_{n-k}$. Therefore,

$$p_A(\lambda) = (-1)^n \sum_{k=0}^{n} (-1)^k \lambda^k a_{n-k}.$$

Replacing k by $n - k$ as the summation index yields

$$p_A(\lambda) = (-1)^n \sum_{k=0}^{n} (-1)^{n-k} \lambda^{n-k} a_k = \sum_{k=0}^{n} (-1)^k a_k \lambda^{n-k}.$$

Definition

Two matrices $A, B \in \mathbb{C}^{n \times n}$ are *similar* if there exists an invertible matrix $P \in \mathbb{C}^{n \times n}$ such that $B = PAP^{-1}$. This is denoted by $A \sim B$.
If there exists a unitary matrix U such that $B = UAU^{-1}$, then A is *unitarily similar* to B. This is denoted by $A \sim_u B$.
The matrices A, B are *congruent* if $B = SAS^{\mathsf{H}}$ for some non-singular matrix S. This is denoted by $A \approx B$. If $A, B \in \mathbb{R}^{n \times n}$, we say that they are *t-congruent* if $B = SAS'$ for some invertible matrix S; this is denoted by $A \approx_t B$.

Similar matrices have the same characteristic polynomial. Indeed, suppose that $B = PAP^{-1}$. We have

$$\begin{aligned} p_B(\lambda) &= \det(\lambda I_n - B) = \det(\lambda I_n - PAP^{-1}) \\ &= \det(\lambda P I_n P^{-1} - PAP^{-1}) = \det(P(\lambda I_n - A)P^{-1}) \\ &= \det(P)\det(\lambda I_n - A)\det(P^{-1}) = \det(\lambda I_n - A) = p_A(\lambda), \end{aligned}$$

because $\det(P)\det(P^{-1}) = 1$. Thus, similar matrices have the same eigenvalues.

Example

Let A be the matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We have

$$p_A = \det(\lambda I_2 - A) = (\lambda - \cos\theta)^2 + \sin^2\theta = \lambda^2 - 2\lambda\cos\theta + 1.$$

The roots of this polynomial are $\lambda_1 = \cos\theta + i\sin\theta$ and $\lambda_2 = \cos\theta - i\sin\theta$, so they are complex numbers.
We regard A as a complex matrix with real entries. If we were to consider A as a real matrix, we would not be able to find real eigenvalues for A unless θ were equal to 0.

Definition

The *algebraic multiplicity* of the eigenvalue λ of a matrix $A \in \mathbb{C}^{n \times n}$ is the multiplicity of λ as a root of the characteristic polynomial p_A of A.
The algebraic multiplicity of λ is denoted by $\text{algm}(A, \lambda)$. If $\text{algm}(A, \lambda) = 1$ we say that λ is a *simple eigenvalue*.

Example

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

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}$$

The characteristic polynomial of A is

$$p_A(\lambda) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda - 2 \end{vmatrix} = \lambda^2 - 2\lambda + 1.$$

Therefore, A has the eigenvalue 1 with $\mathrm{algm}(A, 1) = 2$.

Theorem

The eigenvalues of Hermitian complex matrices are real numbers.

Proof.

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let λ be an eigenvalue of A. We have $A\mathbf{x} = \lambda\mathbf{x}$ for some $\mathbf{x} \in \mathbb{C}^n - \{\mathbf{0}_n\}$, so $\mathbf{x}^{\mathsf{H}} A^{\mathsf{H}} = \overline{\lambda}\mathbf{x}^{\mathsf{H}}$. Since $A^{\mathsf{H}} = A$, we have

$$\lambda\mathbf{x}^{\mathsf{H}}\mathbf{x} = \mathbf{x}^{\mathsf{H}}A\mathbf{x} = \mathbf{x}^{\mathsf{H}}A^{\mathsf{H}}\mathbf{x} = \overline{\lambda}\mathbf{x}^{\mathsf{H}}\mathbf{x}.$$

Since $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{x}^{\mathsf{H}}\mathbf{x} \neq 0$, it follows that $\overline{\lambda} = \lambda$. Thus, λ is a real number. □

Corollary

The eigenvalues of symmetric real matrices are real numbers.

Theorem

The eigenvectors of a complex Hermitian matrix corresponding to distinct eigenvalues are orthogonal to each other.

Proof: Let $(\lambda, \mathbf{u})$ and $(\mu, \mathbf{v})$ be two eigenpairs of the Hermitian matrix $A \in \mathbb{C}^{n\times n}$, where $\lambda \neq \mu$. Since A is Hermitian, $\lambda, \mu \in \mathbb{R}$. Since $A\mathbf{u} = \lambda\mathbf{u}$ we have $\mathbf{v}^{\mathsf{H}}A\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$. The last equality can be written as $(A\mathbf{v})^{\mathsf{H}}\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$, or as $\mu\mathbf{v}^{\mathsf{H}}\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$. Since $\mu \neq \lambda$, $\mathbf{v}^{\mathsf{H}}\mathbf{u} = 0$, so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

UMASS BOSTON

Corollary

The eigenvectors of a Hermitian matrix corresponding to distinct eigenvalues form a linearly independent set.

Theorem

(Schur's Triangularization Theorem) *Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. There exists an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A \sim_u T$. The diagonal elements of T are the eigenvalues of A; moreover, each eigenvalue λ of A occurs in the sequence of diagonal elements of T a number of $\mathrm{algm}(A, \lambda)$ times. The columns of U are unit eigenvectors of A.*

Proof

The argument is by induction on n. The base case, $n = 1$, is immediate. Suppose that the statement holds for matrices in $\mathbb{C}^{(n-1)\times(n-1)}$ and let $A \in \mathbb{C}^{n\times n}$. If $(\lambda, \mathbf{x})$ is an eigenpair of A with $\| \mathbf{x} \|_2 = 1$, let $H_\mathbf{v}$ be a Householder matrix that transforms $\mathbf{x}$ into $\mathbf{e}_1$. Since we also have $H_\mathbf{v}\mathbf{e}_1 = \mathbf{x}$, $\mathbf{x}$ is the first column of $H_\mathbf{v}$ and we can write $H_\mathbf{v} = (\mathbf{x}\ K)$, where $K \in \mathbb{C}^{n\times(n-1)}$. Consequently,

$$H_\mathbf{v} A H_\mathbf{v} = H_\mathbf{v} A(\mathbf{x}\ K) = H_\mathbf{v}(\lambda\mathbf{x}\ H_\mathbf{v} A K) = (\lambda\mathbf{e}_1\ H_\mathbf{v} A K).$$

Proof (cont'd)

Since $H_{\mathbf{v}}$ is Hermitian and $H_{\mathbf{v}} = (\mathbf{x}\ K)$, it follows that

$$H_{\mathbf{v}}^{\mathsf{H}} = \begin{pmatrix} \mathbf{x}^{\mathsf{H}} \\ K^{\mathsf{H}} \end{pmatrix} = H_{\mathbf{v}}.$$

Therefore,

$$H_{\mathbf{v}} A H_{\mathbf{v}} = \begin{pmatrix} \lambda & \mathbf{x}^{\mathsf{H}} A K \\ \mathbf{0}_{n-1} & K^{\mathsf{H}} A K \end{pmatrix}.$$

Proof (cont'd)

Since $K^{\mathsf{H}}AK \in \mathbb{C}^{(n-1)\times(n-1)}$, by the inductive hypothesis, there exists a unitary matrix W and an upper triangular matrix S such that $W^{\mathsf{H}}(K^{\mathsf{H}}AK)W = S$. Note that the matrix

$$U = H_{\mathbf{v}} \begin{pmatrix} 1 & \mathbf{0}'_{n-1} \\ \mathbf{0}_{n-1} & W \end{pmatrix}$$

is unitary and

$$U^{\mathsf{H}}AU^{\mathsf{H}} = \begin{pmatrix} \lambda & x^{\mathsf{H}}AKW \\ \mathbf{0}_{n-1} & W^{\mathsf{H}}K^{\mathsf{H}}AKW \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{x}^{\mathsf{H}}AKW \\ \mathbf{0}_{n-1} & S \end{pmatrix}.$$

The last matrix is clearly upper triangular.

Proof (cont'd)

Since $A \sim_u T$, A and T have the same characteristic polynomials and, therefore, the same eigenvalues, with identical multiplicities.
The factorization of A can be written as $A = UDU^{\mathsf{H}}$ because $U^{-1} = U^{\mathsf{H}}$.
Since $AU = UD$, each column $\mathbf{u}_i$ of U is an eigenvector of A that corresponds to the eigenvalue λ_i for $1 \leqslant i \leqslant n$.

Corollary

Let $A \in \mathbb{C}^{n\times n}$ and let f be a polynomial. If spec$(A) = \{\lambda_1, \ldots, \lambda_n\}$ (including multiplicities), then spec$(f(A)) = \{f(\lambda_1), \ldots, f(\lambda_n)\}$.

Proof: By Schur's Triangularization Theorem there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A = UTU^{\mathsf{H}}$ and the diagonal elements of T are the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$. Therefore $f(A) = Uf(T)U^{\mathsf{H}}$ and the diagonal elements of $f(T)$ are $f(\lambda_1), , \ldots, f(\lambda_m)$. Since $f(A) \sim_u f(T)$, we obtain the desired conclusion because two similar matrices have the same eigenvalues with the same algebraic multiplicities.

Definition

A matrix $A \in \mathbb{C}^{n\times n}$ is *diagonalizable* (unitarily diagonalizable) if there exists a diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ such that $A \sim D$ ($A \sim_u D$).

Theorem

A matrix $A \in \mathbb{C}^{n\times n}$ *is diagonalizable if and only if there exists a linearly independent set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *of* n *eigenvectors of* A*.*

Proof

Let $A \in \mathbb{C}^{n \times n}$ such that there exists a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of n eigenvectors of A that is linearly independent and let P be the matrix $(\mathbf{v}_1\ \mathbf{v}_2\ \ \cdots\ \ \mathbf{v}_n)$ that is clearly invertible. We have:

$$\begin{aligned} P^{-1}AP &= P^{-1}(A\mathbf{v}_1\ A\mathbf{v}_2\ \ \cdots\ \ A\mathbf{v}_n) = P^{-1}(\lambda_1\mathbf{v}_1\ \lambda_2\mathbf{v}_2\ \ \cdots\ \ \lambda_n\mathbf{v}_n) \\ &= P^{-1}P\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \end{aligned}$$

Proof (cont'd)

Therefore, we have $A = PDP^{-1}$, where

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

so $A \sim D$.
Conversely, suppose that A is diagonalizable, so $AP = PD$, where D is a diagonal matrix and P is an invertible matrix, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of the matrix P. We have $A\mathbf{v}_i = d_{ii}\mathbf{v}_i$ for $1 \leqslant i \leqslant n$, so each $\mathbf{v}_i$ is an eigenvector of A. Since P is invertible, its columns are linear independent.

Corollary

If $A \in \mathbb{C}^{n \times n}$ is diagonalizable then the columns of any matrix P such that $D = P^{-1}AP$ is a diagonal matrix are eigenvectors of A. Furthermore, the diagonal entries of D are the eigenvalues that correspond to the columns of P.

Corollary

A matrix $A \in \mathbb{C}^{n \times n}$ *is unitarily diagonalizable if and only if there exists a set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *of* n *orthonormal eigenvectors of* A*.*

Theorem

Let A be a Hermitian matrix, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its eigenvalues having the orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, respectively.
Define the subspace $M = \langle \mathbf{u}_p, \ldots, \mathbf{u}_q \rangle$, where $1 \leqslant p \leqslant q \leqslant n$. If $\mathbf{x} \in M$ and $\| \mathbf{x} \|_2 = 1$, we have $\lambda_q \leqslant \mathbf{x}^H A \mathbf{x} \leqslant \lambda_p$.

Proof

If $\mathbf{x}$ is a unit vector in M, then $\mathbf{x} = a_p\mathbf{u}_p + \cdots + a_q\mathbf{u}_q$, so $\mathbf{x}^{\mathsf{H}}\mathbf{u}_i = \overline{a_i}$ for $p \leqslant i \leqslant q$. Since $\parallel \mathbf{x} \parallel_2 = 1$, we have $|a_p|^2 + \cdots + |a_q|^2 = 1$. This allows us to write:

$$\begin{aligned}
\mathbf{x}^{\mathsf{H}} A\mathbf{x} &= \mathbf{x}^{\mathsf{H}}(a_p A\mathbf{u}_p + \cdots + a_q A\mathbf{u}_q) \\
&= \mathbf{x}^{\mathsf{H}}(a_p\lambda_p\mathbf{u}_p + \cdots + a_q\lambda_q\mathbf{u}_q) \\
&= \mathbf{x}^{\mathsf{H}}(|a_p|^2\lambda_p + \cdots + |a_q|^2\lambda_q).
\end{aligned}$$

Since $|a_p|^2 + \cdots + |a_q|^2 = 1$, the desired inequalities follow immediately.

Corollary

Let A be a Hermitian matrix, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its eigenvalues having the orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, respectively. The following statements hold for a unit vector $\mathbf{x}$:

- *if $\mathbf{x} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_i \rangle$, then $\mathbf{x}^H A \mathbf{x} \geqslant \lambda_i$;*
- *if $\mathbf{x} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_{i-1} \rangle^{\perp}$, then $\mathbf{x}^H A \mathbf{x} \leqslant \lambda_i$.*

UMASS BOSTON

Theorem

(Rayleigh-Ritz Theorem) *Let A be a Hermitian matrix and let* $(\lambda_1, \mathbf{u}_1), \ldots, (\lambda_n, \mathbf{u}_n)$ *be the eigenpairs of A, where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. If $\mathbf{x}$ is a unit vector, we have $\lambda_n \leqslant \mathbf{x}^H A\mathbf{x} \leqslant \lambda_1$.*

Proof.

This statement follows by observing that the subspace generated by $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is the entire space $\mathbb{C}^n$. □

The Courant-Fisher Theorem

Let $\mathcal{S}_p^n$ be the collection of p-dimensional subspaces of $\mathbb{C}^n$. Note that $\mathcal{S}_0^n = \{\{\mathbf{0}_n\}\}$ and $\mathcal{S}_n^n = \{\mathbb{C}^n\}$.

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. We have

$$\begin{aligned}\lambda_k &= \max_{U\in\mathcal{S}_k^n} \min\{\mathbf{x}^H A\mathbf{x} \mid \mathbf{x} \in U \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\} \\ &= \min_{U\in\mathcal{S}_{n-k+1}^n} \max\{\mathbf{x}^H A\mathbf{x} \mid \mathbf{x} \in U \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\}.\end{aligned}$$

Proof

Let $A = V^{\mathsf{H}}\mathrm{diag}(\lambda_1, \ldots, \lambda_n)V$ be the factorization of A, where $V = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n)$ is a unitary matrix.
If

$$U = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle \in \mathcal{S}_k^n \text{ and } W = \langle \mathbf{u}_k, \ldots, \mathbf{u}_n \rangle \in \mathcal{S}_{n-k+1}^n,$$

then there is a non-zero vector $\mathbf{x} \in U \cap W$ because $\dim(U) + \dim(W) = n + 1$; we can assume that $\| \mathbf{x} \|_2 = 1$.
We have $\lambda_k \geqslant \mathbf{x}^{\mathsf{H}} A\mathbf{x}$, and, therefore, for any $U \in \mathcal{S}_k^n$, $\lambda_k \geqslant \min\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}$. This implies

$$\lambda_k \geqslant \max_{U \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}.$$

Proof (cont'd)

For a unit vector $\mathbf{x} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle \in \mathcal{S}_k^n$ we have $\mathbf{x}^{\mathsf{H}} A \mathbf{x} \geqslant \lambda_k$ and $\mathbf{u}_k^{\mathsf{H}} A \mathbf{u}_k = \lambda_k$. Therefore, for $U = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle \in \mathcal{S}_k^n$ we have $\min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U, \| \mathbf{x} \|_2 = 1\} \geqslant \lambda_k$, so $\max_{U \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U, \| \mathbf{x} \|_2 = 1\} \geqslant \lambda_k$. The inequalities proved above yield

$$\lambda_k = \max_{U \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}.$$

Proof (cont'd)

For the second equality, let $U \in \mathcal{S}^n_{n-k+1}$. If $W = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle$, there is a non-zero unit vector $\mathbf{x} \in U \cap W$ because $\dim(U) + \dim(W) \geqslant n + 1$. We have $\mathbf{x}^{\mathsf{H}} A \mathbf{x} \leqslant \lambda_k$. Therefore, for any $U \in \mathcal{S}^n_{n-k+1}$,
$\lambda_k \geqslant \max\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}$. This implies
$\lambda_k \geqslant \min_{U \in \mathcal{S}^n_{n-k+1}} \max\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}$.

Proof (cont'd)

For a unit vector $\mathbf{x} \in \langle \mathbf{u}_k, \ldots, \mathbf{u}_n \rangle \in \mathcal{S}^n_{n-k+1}$ we have $\lambda_k \leqslant \mathbf{x}^{\mathsf{H}} A \mathbf{x}$ and $\lambda_k = \mathbf{u}_k^{\mathsf{H}} A \mathbf{u}_k$. Thus, $\lambda_k \leqslant \max\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}$. Consequently, $\lambda_k \leqslant \min_{U \in \mathcal{S}^n_{n-k+1}} \max\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \| \mathbf{x} \|_2 = 1\}$, which completes the proof of the second equality of the theorem.

An equivalent formulation of Courant-Fisher Theorem is given next.

Theorem

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. We have

$$\begin{aligned}\lambda_k &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}} \min\{\mathbf{x}^H A\mathbf{x} \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{n-k} \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\} \\ &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \max\{\mathbf{x}^H A\mathbf{x} \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{k-1} \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\}.\end{aligned}$$

Proof: The equalities of the Theorem follow from the Courant-Fisher theorem taking into account that if $U \in \mathcal{S}_k^n$, then $U^\perp = \langle \mathbf{w}_1, \ldots, \mathbf{w}_{n-k} \rangle$ for some vectors $\mathbf{w}_1, \ldots, \mathbf{w}n - k$, and if $U \in \mathcal{S}_{n-k+1}^n$, then $U = \langle \mathbf{w}_1, \ldots \mathbf{w}_{k-1} \rangle$ for some vectors $\mathbf{w}_1, \ldots, \mathbf{w}k - 1$ in $\mathbb{C}^n$.

Ky Fan's Theorem

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix such that spec$(A) = \{\lambda_1, \ldots, \lambda_n\}$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Also, let $V \in \mathbb{C}^{n\times n}$ be a matrix, $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ whose set of columns constitutes an orthonormal set of eigenvectors of A. For every $q \in \mathbb{N}$ such that $1 \leqslant q \leqslant n$, the sums

$$\sum_{i=1}^{q} \lambda_i = \lambda_1 + \cdots + \lambda_q$$

and

$$\sum_{i=1}^{q} \lambda_{n+1-i} = \lambda_n + \lambda_{n-1} + \cdots + \lambda_{n-(q-1)}$$

are the maximum and minimum of $\sum_{j=1}^{q} \mathbf{x}_j^H A \mathbf{x}_j$, respectively, where $\{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ is an orthonormal set of vectors in $\mathbb{C}^n$. The maximum (minimum) is achieved when $\mathbf{x}_1, \ldots, \mathbf{x}_q$ are the first (last) columns of V.

Proof

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an orthonormal set of eigenvectors of A and let $\mathbf{x}_i = \sum_{k=1}^{n} b_{ki}\mathbf{v}_k$ be the expression of $\mathbf{x}_i$ using the columns of V as a basis for $1 \leqslant i \leqslant n$. Since each $\mathbf{x}_i$ is a unit vector we have

$$\parallel \mathbf{x}_i \parallel^2 = \mathbf{x}_i^{\mathsf{H}}\mathbf{x}_i = \sum_{k=1}^{n} |b_{ki}|^2 = 1$$

for $1 \leqslant i \leqslant n$. Also, note that

$$\mathbf{x}_i^{\mathsf{H}}\mathbf{v}_r = \left(\sum_{k=1}^{n} \overline{b_{ki}}\mathbf{v}_k^{\mathsf{H}}\right)\mathbf{v}_r = \overline{b_{ri}},$$

due to the orthonormality of the set of columns of V.

Proof (cont'd)

We have

$$\begin{aligned}
\mathbf{x}_i^{\mathrm{H}} A \mathbf{x}_i &= \mathbf{x}_i^{\mathrm{H}} A \sum_{k=1}^{n} b_{ki} \mathbf{v}_k = \sum_{k=1}^{n} b_{ki} \mathbf{x}_i^{\mathrm{H}} A \mathbf{v}_k \\
&= \sum_{k=1}^{n} b_{ki} \mathbf{x}_i^{\mathrm{H}} \lambda_k \mathbf{v}_k = \sum_{k=1}^{n} \lambda_k b_{ki} \overline{b_{ki}} = \sum_{k=1}^{n} |b_{ki}|^2 \lambda_k \\
&= \lambda_q \sum_{k=1}^{n} |b_{ki}|^2 + \sum_{k=1}^{q} (\lambda_k - \lambda_q) |b_{ki}|^2 + \sum_{k=q+1}^{n} (\lambda_k - \lambda_q) |b_{ki}|^2 \\
&\leqslant \lambda_q + \sum_{k=1}^{q} (\lambda_k - \lambda_q) |b_{ki}|^2.
\end{aligned}$$

The last inequality implies

$$\sum_{i=1}^{q} \mathbf{x}_i^{\mathrm{H}} A \mathbf{x}_i \leqslant q \lambda_q + \sum_{i=1}^{q} \sum_{k=1}^{q} (\lambda_k - \lambda_q) |b_{ki}|^2.$$

Proof (cont'd)

Therefore,

$$\sum_{i=1}^{q} \lambda_i - \sum_{i=1}^{q} \mathbf{x}_i^{\mathsf{H}} A \mathbf{x}_i \geqslant \sum_{i=1}^{q} (\lambda_i - \lambda_q) \left(1 - \sum_{k=1}^{q} |b_{ki}|^2\right). \tag{2}$$

We have $\sum_{k=1}^{q} |b_{ik}|^2 \leqslant \| \mathbf{x}_i \|^2 = 1$, so

$$\sum_{i=1}^{q} (\lambda_i - \lambda_q) \left(1 - \sum_{k=1}^{q} |b_{ki}|^2\right) \geqslant 0.$$

The left member of Inequality 2 becomes 0 when $\mathbf{x}_i = \mathbf{v}_i$, so $\sum_{i=1}^{q} \mathbf{x}_i^{\mathsf{H}} A \mathbf{x}_i \leqslant \sum_{i=1}^{q} \lambda_i$. The maximum of $\sum_{i=1}^{q} \mathbf{x}_i^{\mathsf{H}} A \mathbf{x}_i$ is obtained when $\mathbf{x}_i = \mathbf{v}_i$ for $1 \leqslant i \leqslant q$, that is, when X consists of the first q columns of V. The argument for the minimum is similar.

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. If A is positive semidefinite, then all its eigenvalues are non-negative; if A is positive definite then its eigenvalues are positive.

Proof.

Since A is Hermitian all its eigenvalues are real numbers. Suppose that A is positive semidefinite, that is, $\mathbf{x}^{\mathsf{H}} A\mathbf{x} \geqslant 0$ for $\mathbf{x} \in \mathbb{C}^n$. If $\lambda \in \mathrm{spec}(A)$, then $A\mathbf{v} = \lambda\mathbf{v}$ for some eigenvector $\mathbf{v} \neq \mathbf{0}$. The positive semi-definiteness of A implies $\mathbf{v}^{\mathsf{H}} A\mathbf{v} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{v} = \lambda \parallel \mathbf{v} \parallel_2^2 \geqslant 0$, which implies $\lambda \geqslant 0$. It is easy to see that if A is positive definite, then $\lambda > 0$. $\square$

Theorem

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. If A is positive semidefinite, then all its principal minors are non-negative real numbers. If A is positive definite then all its principal minors are positive real numbers.

Proof.

Since A is positive semidefinite, every sub-matrix $A\begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix}$ is a Hermitian positive semidefinite matrix by Theorem **??**, so every principal minor is a non-negative real number. The second part of the theorem is proven similarly. $\square$

Corollary

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. The following statements are equivalent.

- *A is positive semidefinite;*
- *all eigenvalues of A are non-negative numbers;*
- *there exists a Hermitian matrix $C \in \mathbb{C}^{n \times n}$ such that $C^2 = A$;*
- *A is the Gram matrix of a sequence of vectors, that is, $A = B^H B$ for some $B \in \mathbb{C}^{n \times n}$.*

Definition

Let $A \in \mathbb{C}^{m \times n}$ be a matrix. A *singular triplet* of A is a triplet $(\sigma, \mathbf{u}, \mathbf{v})$ such that $\sigma \in \mathbb{R}_{>0}$, $\mathbf{u} \in \mathbb{C}^n$, $\mathbf{v} \in \mathbb{C}^m$, $A\mathbf{u} = \sigma\mathbf{v}$ and $A^{\mathsf{H}}\mathbf{v} = \sigma\mathbf{u}$.
The number σ is a *singular value* of A, $\mathbf{u}$ is a *left singular vector* and $\mathbf{v}$ is a *right singular vector*.

For a singular triplet $(\sigma, \mathbf{u}, \mathbf{v})$ of A we have

$$A^{\mathsf{H}} A\mathbf{u} = \sigma A^{\mathsf{H}}\mathbf{v} = \sigma^2\mathbf{u} \text{ and } AA^{\mathsf{H}}\mathbf{v} = \sigma A\mathbf{u} = \sigma^2\mathbf{v}.$$

Therefore, σ^2 is both an eigenvalue of AA^{H} and an eigenvalue of $A^{\mathsf{H}}A$.

Example

Let A be the real matrix

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \cos\beta & \sin\beta \end{pmatrix}.$$

We have $\det(A) = \sin(\beta - \alpha)$, so the eigenvalues of $A'A$ are the roots of the equation $\lambda^2 - 2\lambda + \sin^2(\beta - \alpha) = 0$, that is, $\lambda_1 = 1 + \cos(\beta - \alpha)$ and $\lambda_2 = 1 - \cos(\beta - \alpha)$. Therefore, the singular values of A are $\sigma_1 = \sqrt{2}\left|\cos\frac{\beta-\alpha}{2}\right|$ and $\sigma_2 = \sqrt{2}\left|\sin\frac{\beta-\alpha}{2}\right|$.
It is easy to see that a unit left singular vector that corresponds to the eigenvalue $1 + \cos(\beta - \alpha)$ is

$$\mathbf{u} = \begin{pmatrix} \cos\frac{\alpha+\beta}{2} \\ \sin\frac{\alpha+\beta}{2} \end{pmatrix},$$

which corresponds to the average direction of the rows of A.

- The eigenvalues of a positive semi-definite matrix are non-negative numbers. Since both AA^{H} and $A^{\mathsf{H}}A$ are positive semi-definite matrices for $A \in \mathbb{C}^{m \times n}$, the spectra of these matrices consist of non-negative numbers $\lambda_1, \ldots, \lambda_n$.
- AA^{H} and $A^{\mathsf{H}}A$ have the same rank r and therefore, the same number r of non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$.
- The singular values of A have the form $\sqrt{\lambda_1} \geqslant \cdots \geqslant \sqrt{\lambda_r}$.

Notation: $\sigma_i = \sqrt{\lambda_i}$ for $1 \leqslant i \leqslant r$ and will assume that $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$.

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a matrix having the singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_n$. If λ is an eigenvalue value of A, then $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$.

Proof.

Let $\mathbf{u}$ be an unit eigenvector for the eigenvalue λ. Since $A\mathbf{u} = \lambda\mathbf{u}$ it follows that $(A^{\mathsf{H}}A\mathbf{u}, \mathbf{u}) = (A\mathbf{u}, A\mathbf{u}) = \overline{\lambda}\lambda(\mathbf{u}, \mathbf{u}) = \overline{\lambda}\lambda = |\lambda|^2$. The matrix $A^{\mathsf{H}}A$ is Hermitian and its largest and smallest eigenvalues are σ_1^2 and σ_n^2, respectively. Thus, $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$. □

The SVD Theorem

Theorem

If $A \in \mathbb{C}^{m \times n}$ *is a matrix and* $rank(A) = r$, *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, and* $D = diag(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{m \times n}$, *where* $\sigma_1 \geqslant \ldots \geqslant \sigma_r$ *are real positive numbers.*

Proof

The square matrix $A^{\mathsf{H}}A \in \mathbb{C}^{n\times n}$ has the same rank r 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 $1 \leqslant i \leqslant r$. Define $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r \ \mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$ by completing the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to an orthogonal basis

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

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 $V = (V_1 \ V_2)$.

Proof (cont'd)

The equalities involving the eigenvectors can now be written as $A^{\mathsf{H}}AV_1 = V_1E^2$, where $E = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.
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.

Corollary

Let $A \in \mathbb{C}^{m \times n}$ be a matrix such that rank$(A) = r$. If $\sigma_1 \geqslant \ldots \geqslant \sigma_r$ are non-zero singular values, then

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^H, \tag{3}$$

where $(\sigma_i, \mathbf{u}_i, \mathbf{v}_i)$ are singular triplets of A for $1 \leqslant i \leqslant r$.

The value of a unitarily invariant norm of a matrix depends only on its singular values.

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

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

Proof.

This statement follows because the matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary. □

$\|\cdot\|_2$ and $\|\cdot\|_F$ are unitarily invariant. Therefore, the Frobenius norm can be written as

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

and $\|A\|_2 = \sigma_1$.

Theorem

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

Proof.

Suppose that $A \sim_u B$, 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 $\mathbf{v} \in \mathbb{C}^n$ 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 $A \in \mathbb{C}^{m\times n}$ be a matrix such that $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$ is the non-increasing sequence of singular values of A. For $1 \leqslant k \leqslant r$ we have

$$\begin{aligned} \sigma_k &= \min_{\dim(S)=n-k+1} \max\{\parallel A\mathbf{x} \parallel_2 \mid \mathbf{x} \in S \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\} \\ \sigma_k &= \max_{\dim(T)=k} \min\{\parallel A\mathbf{x} \parallel_2 \mid \mathbf{x} \in T \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\}, \end{aligned}$$

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 square root of 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.

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 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 $B(k)\mathbf{x} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \mathbf{x} = \mathbf{0}$. 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 $1 \leqslant i \leqslant k$. Thus, $\mathbf{x} \in \langle \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle^{\perp}$ and, since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent it follows that $\dim(\text{NullSp}(B(k)) = n - k$, which implies $rank(B(k)) = k$ for $1 \leqslant k \leqslant r$. $\square$

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 $r_k = \inf\{\|A - X\|_2 \mid X \in \mathbb{C}^{m \times n}$ *and 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}. Since σ_{k+1} is the largest singular value of $A - B(k)$ we have $\|A - B(k)\|_2 = \sigma_{k+1}$ for $1 \leqslant k \leqslant r$.

Proof (cont'd)

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 unit non-zero vector $\mathbf{x}$.
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

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

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

It is interesting to observe that the matrix $B(k)$ provides an optimal approximation of A not only with respect to $\|\!| \cdot |\!\|_2$ but also relative to the Frobenius norm.

Theorem

$B(k)$ is the best approximation of A among matrices of rank no larger than k in the sense of the Frobenius norm.

Proof

Note that $\| A - B(k) \|_F^2 = \| A \|_F^2 - \sum_{i=1}^k \sigma_i^2$. Let X be a matrix of rank k, which can be written as $X = \sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i^{\mathsf{H}}$. Without loss of generality we may assume that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are orthonormal. If this is not the case, we can use the Gram-Schmidt algorithm to express then as linear combinations of orthonormal vectors, replace these expressions in $\sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i^{\mathsf{H}}$ and rearrange the terms. Now, the Frobenius norm of $A - X$ can be written as

$$\begin{aligned}
\| A - X \|_F^2 &= \mathit{trace}\left(\left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}^{\mathsf{H}}\right)^{\mathsf{H}} \left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}^{\mathsf{H}}\right)\right) \\
&= \mathit{trace}\left(A^{\mathsf{H}} A + \sum_{i=1}^k (\mathbf{y}_i - A^{\mathsf{H}} \mathbf{x}_i)(\mathbf{y}_i - A^{\mathsf{H}} \mathbf{x}_i)^{\mathsf{H}} - \sum_{i=1}^k A^{\mathsf{H}} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}} A\right)
\end{aligned}$$

Taking into account that $\sum_{i=1}^{k}(\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)(\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)^{\mathsf{H}}$ is a real non-negative number and that $\sum_{i=1}^{k} A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}A = \parallel A\mathbf{x}_i \parallel_F^2$ we have

$$\parallel A - X \parallel_F^2 \quad \geq \quad trace\left(A^{\mathsf{H}}A - \sum_{i=1}^{k} A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}A\right) = \parallel A \parallel_F^2 - trace\left(\sum_{i=1}^{k} A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}A\right)$$

Let $A = U\text{diag}(\sigma_1, \ldots, \sigma_n)V^{\mathsf{H}}$ be the singular value decomposition of A. If $V = (V_1\ V_2)$, where V_1 has k columns $\mathbf{v}_1, \ldots, \mathbf{v}_k$, $D_1 = \text{diag}(\sigma_1, \ldots, \sigma_k)$ and $D_2 = \text{diag}(\sigma_{k+1}, \ldots, \sigma_n)$, then

$$\begin{aligned} A^{\mathsf{H}}A &= VD^{\mathsf{H}}U^{\mathsf{H}}UDV^{\mathsf{H}} = (V_1\ V_2)\begin{pmatrix} D_1^2 & O \\ O & D_2^2 \end{pmatrix}\begin{pmatrix} V_1^{\mathsf{H}} \\ V_2^{\mathsf{H}} \end{pmatrix} \\ &= V_1 D_1^2 V_1^{\mathsf{H}} + V_2 D_2^2 V_2^{\mathsf{H}}. \end{aligned}$$

and $A^{\mathsf{H}}A = VD^2V^{\mathsf{H}}$.

Proof (cont'd)

These equalities allow us to write:

$$
\begin{aligned}
\parallel A\mathbf{x}_i \parallel_F^2 &= trace(\mathbf{x}_i^{\mathsf{H}} A^{\mathsf{H}} A\mathbf{x}_i) \\
&= trace\left(\mathbf{x}_i^{\mathsf{H}} V_1 D_1^2 V_1^{\mathsf{H}}\mathbf{x}_i + \mathbf{x}_i^{\mathsf{H}} V_2 D_2^2 V_2^{\mathsf{H}}\mathbf{x}_i\right) \\
&= \parallel D_1 V_1^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 + \parallel D_2 V_2^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 \\
&= \sigma_k^2 + \left(\parallel D_1 V_1^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}}\mathbf{x}_i \parallel_F^2\right) \\
&\quad - \left(\sigma_k^2 \parallel V_2^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 - \parallel D_2 V_2^{\mathsf{H}}\mathbf{x}_i \parallel_F^2)\right) - \sigma_k^2(1 - \parallel V^{\mathsf{H}}\mathbf{x}_i \parallel).
\end{aligned}
$$

Since $\parallel V^{\mathsf{H}}\mathbf{x}_i \parallel_F^1 = 1$ (because $\mathbf{x}_i$ is an unit vector and V is an unitary matrix) and $\sigma_k^2 \parallel V_2^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 - \parallel D_2 V_2^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 \geqslant 0$, it follows that

$$\parallel A\mathbf{x}_i \parallel_F^2 \leqslant \sigma_k^2 + \left(\parallel D_1 V_1^{\mathsf{H}}\mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}}\mathbf{x}_i \parallel_F^2\right).$$

Proof (cont'd)

Consequently,

$$
\begin{aligned}
\sum_{i=1}^{k} \parallel A\mathbf{x}_i \parallel_F^2 &\leqslant k\sigma_k^2 + \sum_{i=1}^{k} \left(\parallel D_1 V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2\right) \\
&= k\sigma_k^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} (\sigma_j^2 - \sigma_k^2) |\mathbf{v}_j^{\mathsf{H}} \mathbf{x}_i|^2 \\
&= \sum_{j=1}^{k} \left(\sigma_k^2 + (\sigma_j^2 - \sigma_k^2) \sum_{i=1}^{k} |\mathbf{v}_j \mathbf{x}_i|^2 \right) \\
&\leqslant \sum_{j=1}^{k} (\sigma_k^2 + (\sigma_j^2 - \sigma_k^2)) = \sum_{j=1}^{k} \sigma_j^2,
\end{aligned}
$$

which concludes the argument.