Chapter 6

Eigenvalues

6.1 Eigenvalues and Eigenvectors

Definition 6.1.1 (Eigenvalue and Eigenvector). *Let A be an $n \times n$ matrix. If there exists a scalar λ and a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ such that*

$$A\mathbf{x} = \lambda\mathbf{x},$$

*then we call λ an **eigenvalue** of A and $\mathbf{x}$ an **eigenvector** of A corresponding to λ.*

Definition 6.1.2 (Characteristic Polynomial). *The polynomial $p(\lambda) = det(\lambda I - A)$ is called the **characteristic polynomial** of A.*

For an $n \times n$ matrix A, $p(\lambda)$ is a polynomial of degree n and the eigenvalues of A are the zeros of $p(\lambda) = 0$. Thus, A will have n eigenvalues, some of which may be repeated and some of which may be complex numbers.

Theorem 6.1.3. *Let A be a square matrix and λ be a scalar. The following statements are equivalent:*

(a) The scalar λ is an eigenvalue of A.

(b) The system $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

(c) The nullsapce $N(A-\lambda I) \neq \{\mathbf{0}\}$.

(d) The matrix $A-\lambda I$ *is singular.*

(e) $\det(A-\lambda I)=0$.

Remark 6.1.4. *Some useful facts:*

1. *A scalar* λ *is an eigenvalue of* A *if and only if* λ *is a root of* $\det(A-\lambda I)=0$.
2. *Any nonzero vector* $\mathbf{x} \in N(A-\lambda I)$ *is an eigenvector of* A *related to eigenvalue* λ.
3. *In particular, a nonzero vector* $\mathbf{v} \in N(A)$ *is an eigenvector of* A *related to* $\lambda=0$.

Example 6.1.5. Find the eigenvalues and the corresponding eigenvectors of A

$$A=\begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}.$$

Solution: Solve $\det(A-\lambda I)=\begin{vmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{vmatrix}=-\lambda(\lambda-1)^2=0.$

Thus, the characteristic polynomial has roots $\lambda_1=0$, $\lambda_2=\lambda_3=1$.

For $\lambda_1=0$, the corresponding eigenvectors $\mathbf{v}_1 \in \mathrm{N}(A)$ can be determined by

$$\left(\begin{array}{ccc|c} 2 & -3 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & -3 & 2 & 0 \end{array}\right) \xrightarrow[R_3-R_2\to R_3]{R_1-2R_2\to R_1} \left(\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right) \xrightarrow[R_2+2R_1\to R_2]{R_3+R_1\to R_3} \left(\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Hence, any nonzero multiple of $(1,1,1)^T$ is an eigenvector belonging to λ_1.

For $\lambda_{2,3}=1$, the corresponding eigenvectors $\mathbf{v}_2 \in \mathrm{N}(A-I)$ can be determined by

$$\left(\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \end{array}\right) \xrightarrow[R_3-R_1\to R_3]{R_2-R_1\to R_2} \left(\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \to \mathrm{N}(A-I)=\operatorname{span}\left\{\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

All nonzero vectors of the form $\alpha(3,1,0)^T+\beta(-1,0,1)^T$ are eigenvectors w.r.t. $\lambda_{2,3}$. ထ

Definition 6.1.6 (Eigenspace). *If λ is an eigenvalue of A, then E_λ is called the* ***eigenspace*** *of A corresponding to λ, where*

$$E_\lambda = N(A - \lambda I) = \{\mathbf{0}\} \cup \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \lambda \mathbf{v}, \quad \textit{for some } \mathbf{v} \neq \mathbf{0}\}.$$

Then E_λ is a subspace of $\mathbb{R}^n$.

Example 6.1.7. Find the eigenvalues and the corresponding eigenspaces of A

$$A = \begin{bmatrix} 2 & -2 & 1 \\ 0 & 4 & 1 \\ 0 & 4 & 4 \end{bmatrix}.$$

Solution: The characteristic equation is

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -2 & 1 \\ 0 & 4-\lambda & 1 \\ 0 & 4 & 4-\lambda \end{vmatrix} = (2-\lambda)^2(6-\lambda) = 0.$$

Eigenvalues of A are $\lambda_1 = 6$ and $\lambda_{2,3} = 2$.

For $\lambda_1 = 6$, the eigenspace corresponding to $\lambda_1 = 6$ is $E_6 = \text{span}\{(0,1,2)^T\}$ since

$$\left(\begin{array}{ccc|c} -4 & -2 & 1 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 4 & -2 & 0 \end{array}\right) \xrightarrow[R_3+2R_2 \to R_3]{R_1-R_2 \to R_1} \left(\begin{array}{ccc|c} -4 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

For $\lambda_{2,3} = 2$, the corresponding eigenvectors $\mathbf{v}_2 \in \text{N}(A - 2I)$ can be determined by

$$\left(\begin{array}{ccc|c} 0 & -2 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 4 & 2 & 0 \end{array}\right) \xrightarrow[R_1+R_2 \to R_1]{R_3-2R_2 \to R_3} \left(\begin{array}{ccc|c} 0 & 0 & 2 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \to \text{N}(A - I) = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right\}.$$

The eigenspace corresponding to $\lambda_{2,3}$ is $E_2 = \text{span}\{(1,0,0)^T\}$. ◎

Remark 6.1.8. *Suppose A has an eigenvalue λ corresponding to an eigenvector* $\mathbf{x}$.

(a) For any positive integer k, A^k has λ^k as an eigenvalue belonging to eigenvector $\mathbf{x}$.

(b) If A is nonsingular, then A^{-1} has $\frac{1}{\lambda}$ as an eigenvalue and $\mathbf{x}$ as an eigenvector.

(c) A^T and A have the same eigenvalues.

(d) If B is similar to A, then A and B have the same eigenvalues.

(e) Let β be a constant. Then $\beta\lambda$ is an eigenvalue of βA belonging to eigenvector $\mathbf{x}$.

(f) The matrix A is singular if and only if 0 is an eigenvalue of A

Exercise 6.1.9. Prove properties (a)-(f) in the remark as an exercise. ✓

Theorem 6.1.10 (Eigenvalues of a Triangular Matrix)**.** *Let A be a triangular matrix, then the eigenvalues of A are diagonal elements.*

Proof. Let A be an $n \times n$ matrix. Consider the matrix $(A - \lambda I)$, since A is a triangular matrix, so $(A - \lambda I)$ is a triangular matrix of the same type as A. Then its characteristic equation is

$$\det(A - \lambda I) = (a_{11} - \lambda) \cdots (a_{nn} - \lambda) = \prod_{i=1}^{n} (a_{ii} - \lambda) = 0$$

Thus, eigenvalues of A are a_{ii}'s, the diagonal elements of A. □

Definition 6.1.11 (Algebraic Multiplicity)**.** *Suppose $p(\lambda)$ is the characteristic polynomial of an $n \times n$ matrix. Let*

$$p(\lambda) = \prod_{i=1}^{k} (\lambda_i - \lambda)^{m_i}, \qquad \sum_{i=1}^{k} m_i = n$$

be a factorization of $p(\lambda)$. Then m_i is called the ***algebraic multiplicity*** *of the eigenvalue λ_i for each $i = 1, \ldots, k$.*

Definition 6.1.12 (Geometric Multiplicity)**.** *The **geometric multiplicity** of an eigenvalue λ is the dimension of the eigenspace corresponding to λ.*

Theorem 6.1.13. *The geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity.*

Example 6.1.14. In Example 6.1.5, $\text{almu}(0) = 1$, $\text{gemu}(0) = \dim E_0 = 1$; And $\text{almu}(1) = 2$, $\text{gemu}(2) = \dim E_2 = 2$.
In Example 6.1.7, $\text{almu}(6) = 1$, $\text{gemu}(6) = 1$; And $\text{almu}(2) = 2$, $\text{gemu}(2) = 1$. ⚭

Exercise 6.1.15. Eigenvalues and its algebraic multiplicity by observation:

$$A = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

✓

6.1.1 The Product and Sum of the Eigenvalues

Theorem 6.1.16. *Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues (with possible repetitions) of an $n \times n$ matrix A, then*

1. $\displaystyle\prod_{i=1}^{n} \lambda_i = \lambda_1 \cdot \lambda_2 \cdot \cdots \cdot \lambda_n = \det(A)$,
2. $\displaystyle\sum_{i=1}^{n} \lambda_i = tr(A) := \sum_{i=1}^{n} a_{ii}$.

Proof. Consider the constant term and the first degree term of $p(\lambda)$. □

Example 6.1.17. Find all eigenvalues and eigenvectors of

$$A = \begin{bmatrix} a & a & \cdots & a \\ a & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & a \end{bmatrix}_{n\times n} \quad \text{and} \quad B = \begin{bmatrix} b & a & \cdots & a \\ a & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & b \end{bmatrix}_{n\times n}.$$

Solution: Since $\operatorname{rank}(A) = 1$, $\dim \mathrm{N}(A) = n - 1$ by rank-nullity Theorem. Clearly,

$$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}$$

form a basis for $\mathrm{N}(A)$, which are also eigenvectors of A corresponding to 0. Let λ be the last eigenvalue of A. By Theorem 6.1.16, we have $0 + \cdots + 0 + \lambda = Tr(A) = na$. Then $\lambda = na$, consider

$$\mathrm{N}(A - \lambda I) = \mathrm{N}\begin{pmatrix} -(n-1)a & a & \cdots & a \\ a & -(n-1)a & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & -(n-1)a \end{pmatrix} = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\right\}$$

For matrix B, we find $B = A + (b-a)I$. Let μ be an eigenvalue of A with eigenvector $\mathbf{x}$, i.e. $A\mathbf{x} = \mu\mathbf{x}$, then

$$B\mathbf{x} = [A + (b-a)I]\mathbf{x} = \mu\mathbf{x} + (b-a)\mathbf{x} = (\mu + b - a)\mathbf{x}$$

Thus, $\mu + b - a$ is an eigenvalue of B w.r.t. eigenvector $\mathbf{x}$.

Eigenvalues of B are $(b - a)$ with algebraic multiplicity $n - 1$ and $(\lambda + b - a)$.

Corresponding eigenspaces are subspaces of $\mathbb{R}^n$ and

$$E_{b-a} = \text{span}\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} \right\} \quad \text{and} \quad E_{\lambda+b-a} = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \right\}.$$

∞

A real-valued square matrix may have complex eigenvalues and eigenvectors.

Example 6.1.18. Compute the eigenvalues of A and find bases for the corresponding eigenspaces.

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

Solution: The characteristic equation is

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 5 = 0.$$

The eigenvalues of A are $\lambda_{1,2} = \dfrac{2 \pm \sqrt{4-20}}{2} = 1 \pm 2i$.

$$A - (1+2i)I = \begin{bmatrix} -2i & 2 \\ -2 & -2i \end{bmatrix}$$

It follows that $E_{1+2i} = \text{span}\{(1,i)^T\}$ is the eigenspace corresponding to $\lambda_1 = 1+2i$.

$$A - (1-2i)I = \begin{bmatrix} 2i & 2 \\ -2 & 2i \end{bmatrix}$$

and $\{(1,-i)^T\}$ is a basis for $E_{1-2i} = \mathrm{N}(A - \lambda_2 I)$. ∞

Observations: For a real-valued square matrix A, if there exists a complex eigenvalue $\lambda_1 = a + ib$ of A, then the corresponding eigenvector is also complex $\mathbf{v}_1 = \mathbf{x} + i\mathbf{y}$. Further, $\lambda_2 = a - ib$ and $\mathbf{v}_2 = \mathbf{x} - i\mathbf{y}$ also form an eigenpair of A.

Exercise 6.1.19. Find all eigenvalues and eigenvectors:

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

Answer: $\lambda_1 = 2, \lambda_2 = 1+2i$, $\lambda_3 = 1-2i$. $\mathbf{x}_1 = (5, 2, 1)'$, $\mathbf{x}_2 = (0, 1, i)'$ *and* $\mathbf{x}_3 = (0, 1, -i)'$. ✓

Question: What have you found from the last exercise in the view of a block matrix?

6.3 Diagonalization

Definition 6.3.1. *An $n \times n$ matrix A is **diagonalizable** if there exists a nonsingular matrix P and a diagonal matrix D so that*

$$A = PDP^{-1},$$

that is, A is similar to a diagonal matrix.

Theorem 6.3.2. *An $n \times n$ matrix is diagonalizable if and only if it has n linearly independent eigenvectors.*

Proof. Suppose that the matrix A has n linearly independent eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_n$. Let λ_i be the eigenvalue of A corresponding to $\mathbf{x}_i$ for each i. (Some of the λ_is may be equal.) Let P be the matrix whose jth column vector is $\mathbf{x}_j$ for $j = 1, \cdots, n$. It follows that $A\mathbf{x}_j = \lambda_j \mathbf{x}_j$ is the jth column vector of AP. Thus,

$$\begin{aligned} AP &= (A\mathbf{x}_1, A\mathbf{x}_2, \cdots, A\mathbf{x}_n) \\ &= (\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \cdots, \lambda_n\mathbf{x}_n) \\ &= (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ &= PD \end{aligned}$$

Since P has n linearly independent column vectors, it follows that P is nonsingular and

$$A = PDP^{-1}$$

Conversely, if A is diagonalizable, then there exists a nonsingular matrix P such that $AP = PD$. If $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_n$ are the column vectors of P, then for each j,

$$A\mathbf{x}_j = \lambda_j \mathbf{x}_j \qquad\qquad (\lambda_j = d_{jj})$$

Thus, for each j, λ_j is an eigenvalue of A and $\mathbf{x}_j$ is an eigenvector belonging to λ_j. Since the column vectors of P are linearly independent, it follows that A has n linearly independent eigenvectors. □

Corollary 6.3.3. *Let $A = PDP^{-1}$ be a diagonalizable matrix. Then for each k, the k-th diagonal entry of D is an eigenvalue of A corresponding to the eigenvector as the k-th column of P.*

$$D = \begin{bmatrix} \lambda_{11} & & & \\ & \lambda_{22} & & \\ & & \ddots & \\ & & & \lambda_{nn} \end{bmatrix}, \qquad P = \left[\begin{array}{c|c|c|c} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array}\right].$$

Example 6.3.4. In Example 6.1.5, let $D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $P_1 = \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. Then

$$A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} = P_1 D_1 P_1^{-1}$$

Similarly, if let $D_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, then $P_2 = \begin{bmatrix} -1 & 1 & 3 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. Then

$$A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 3 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & 2 \\ -1 & 3 & -1 \\ 1 & -2 & 1 \end{bmatrix} = P_2 D_2 P_2^{-1}$$

Remarks: The diagonalizing matrix P is not unique. Reordering the columns of a given matrix P or multiplying them by nonzero scalars will produce a new diagonalizing matrix. Accordingly, the diagonal matrix D may change along with matrix P.

6.3.1 Diagonalizable Matrices

Theorem 6.3.5. *If $\lambda_1, \lambda_2, \cdots, \lambda_k$ are distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$, then the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly independent.*

Proof. Consider

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

for some constants $c_1, c_2, \cdots, c_k$. Multiplying A to the left on both sides of the equation, we have

$$A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = A\mathbf{0} = \mathbf{0}$$

□

Corollary 6.3.6. *If an $n \times n$ matrix has n distinct eigenvalues, then it is diagonalizable.*

Theorem 6.3.7. *If A is diagonalizable, then the rank of A equals the number of nonzero eigenvalues of A.*

If an $n \times n$ matrix A has less than n distinct eigenvalues, then A may or may not be diagonalizable.

Corollary 6.3.8. *If $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of A and $B_1, \dots, B_k$ are bases for the corresponding eigenspaces. Then $B = \bigcup_{i=1}^{k} B_i$ is linearly independent.*

Theorem 6.3.9. *A matrix is diagonalizable if and only if for each eigenvalue, the geometric multiplicity equals the algebraic multiplicity. That is,*

$$\mathit{almu}(\lambda_i) = \mathit{gemu}(\lambda_i), \qquad \text{for each } i$$

6.3.2 Defective Matrices**

Definition 6.3.10. *A square matrix is said to be **defective** if it is not diagonalizable.*

A defective matrix A is not diagonalizable because eigenspace E_λ does not provide enough linearly independent eigenvectors. We may extend the eigenspace E_λ to generalized eigenspace K_λ.

Definition 6.3.11 (Generalized Eigenvector and Eigenspace)**.** *Let $A \in \mathbb{C}^{n \times n}$ and λ be an eigenvalue of A. A nonzero vector $\mathbf{v} \in \mathbb{C}^n$ is called a* **generalized eigenvector** *of A corresponding to λ if*

$$(A - \lambda I_n)^p \mathbf{v} = \mathbf{0}$$

for some positive integer p. The **generalized eigenspace** *of A corresponding to λ, denoted K_λ, is defined by*

$$K_\lambda = \{\mathbf{v} \in \mathbb{C}^n \mid (A - \lambda I_n)^p \mathbf{v} = \mathbf{0} \ \text{ for some positive } p\}.$$

Theorem 6.3.12 (Jordan Canonical Form)**.** *Any $n \times n$ matrix A can be factored into a product*

$$A = PJP^{-1},$$

where J is an upper bidiagonal matrix with the eigenvalues of A on its main diagonal and 0's and 1's on the diagonal directly above the main diagonal. Here, J is of **Jordan Canonical Form**. *When A is diagonalizable, J is simply the diagonal matrix.*

6.4 Hermitian Matrices

Let $\mathbb{C}^n$ denote the vector space of all n-tuples of complex numbers. The set $\mathbb{C}$ of all complex numbers will be taken as the field of scalars. We have already seen that a matrix A with real entries may have complex eigenvalues and eigenvectors. In this section, we study matrices with complex entries and look at the complex analogues of symmetric and orthogonal matrices.

If $\alpha = a + bi$ is a complex scalar, the length of α is given by

$$|\alpha| = \sqrt{\bar{\alpha}\alpha} = \sqrt{a^2 + b^2}$$

The length of a vector $\mathbf{z} = (z_1, z_2, \cdots, z_n)^T$ in $\mathbb{C}^n$ is given by

$$\begin{aligned} \|\mathbf{z}\| &= \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2} \\ &= (\bar{z}_1 z_1 + \bar{z}_2 z_2 + \cdots + \bar{z}_n z_n)^{1/2} \\ &= (\bar{\mathbf{z}}^T \mathbf{z})^{1/2} \end{aligned}$$

As a notational convenience, we write $\mathbf{z}^H$ for the transpose of $\bar{\mathbf{z}}$. Thus $\bar{\mathbf{z}}^T = \mathbf{z}^H$ and $\|\mathbf{z}\| = (\mathbf{z}^H \mathbf{z})^{1/2}$. The vector $\mathbf{z}^H$ is said to be the **conjugate transpose** of $\mathbf{z}$.

If $\mathbf{z} = \mathbf{a} + \mathbf{b}i$ is a complex vector, the length of $\mathbf{z}$ is given by

$$\begin{aligned} \|\mathbf{z}\| &= \sqrt{\mathbf{z}^H \mathbf{z}} = \sqrt{(\mathbf{a} - \mathbf{b}i)^T (\mathbf{a} + \mathbf{b}i)} \\ &= (\mathbf{a}^T \mathbf{a} - \mathbf{b}^T \mathbf{a} i + \mathbf{a}^T \mathbf{b} i + \mathbf{b}^T \mathbf{b})^{1/2} \\ &= \sqrt{\mathbf{a}^T \mathbf{a} + \mathbf{b}^T \mathbf{b}} \end{aligned}$$

6.4.1 Complex Inner Products

Definition 6.4.1 (Complex Inner Product)**.** *Let V be a vector space over $\mathbb{C}$. An* ***inner product*** *on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ satisfying the following conditions: For any $\mathbf{z}, \mathbf{w}$ and $\mathbf{u}$ in V, and any $\alpha, \beta \in \mathbb{C}$,*

1. $\langle \mathbf{z}, \mathbf{z} \rangle \geq 0$ *with equality holding only if* $\mathbf{z} = \mathbf{0}$*;*
2. $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}$*;* $\leftarrow$ *Order matters!*
3. $\langle \alpha \mathbf{z} + \beta \mathbf{w}, \mathbf{u} \rangle = \alpha \langle \mathbf{z}, \mathbf{u} \rangle + \beta \langle \mathbf{w}, \mathbf{u} \rangle$.

Exercise 6.4.2 ($\mathbb{C}^n$)**.** Show that an inner product on $\mathbb{C}^n$ can be defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{x} = x_1 \bar{y}_1 + \cdots + x_n \bar{y}_n, \qquad \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$$

✓

Example 6.4.3. Let the inner product on $\mathbb{C}^2$ be defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{x}$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^2$. Compute $\langle \mathbf{z}, \mathbf{w} \rangle$, $\langle \mathbf{w}, \mathbf{y} \rangle$, $\langle \mathbf{y}, \mathbf{w} \rangle$ and the length of each vector for

$$\mathbf{z} = \begin{pmatrix} 5+i \\ 1-3i \end{pmatrix}, \ \mathbf{y} = \begin{pmatrix} 1+i \\ 2-i \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} 2+i \\ -2+3i \end{pmatrix}$$

The vector lengths are $\|\mathbf{z}\| = 6$, $\|\mathbf{y}\| = 2\sqrt{2}$ and $\|\mathbf{w}\| = 3\sqrt{2}$ since

$$\begin{aligned} \mathbf{z}^H \mathbf{z} &= \begin{pmatrix} 5-i & 1+3i \end{pmatrix} \begin{pmatrix} 5+i \\ 1-3i \end{pmatrix} = (25+1) + (1+9) = 36 \\ \mathbf{y}^H \mathbf{y} &= (1,2)(1,2)^T + (1,-1)(1,-1)^T = (1+5) + (1+1) = 8 \\ \mathbf{w}^H \mathbf{w} &= |2+i|^2 + |-2+3i|^2 = 18 \end{aligned}$$

The inner products are

$$\begin{aligned}
\langle \mathbf{z}, \mathbf{w}\rangle = \mathbf{w}^H\mathbf{z} &= \begin{pmatrix} 2-i & -2-3i \end{pmatrix}\begin{pmatrix} 5+i \\ 1-3i \end{pmatrix} = (11-3i) + (-11++3i) = 0 \\
\langle \mathbf{w}, \mathbf{y}\rangle = \mathbf{y}^H\mathbf{w} &= \begin{pmatrix} 1-i & 2+i \end{pmatrix}\begin{pmatrix} 2+i \\ -2+3i \end{pmatrix} = (4-i) + (-7+4i) = -4+3i \\
\langle \mathbf{y}, \mathbf{w}\rangle = \mathbf{w}^H\mathbf{y} &= \begin{pmatrix} 2-i & -2-3i \end{pmatrix}\begin{pmatrix} 1+i \\ 2-i \end{pmatrix} = -4-3i = \overline{\langle \mathbf{w}, \mathbf{y}\rangle}
\end{aligned}$$

ꝏ

6.4.2 Hermitian Matrix and Unitary Matrix

Let $M = A+iB$ be an $m \times n$ matrix with $m_{ij} = a_{ij} + ib_{ij}$, where $A = (a_{ij})$ and $B = (b_{ij})$ have real entries. Define the conjugate of M by

$$\overline{M} = A - iB$$

Thus, $\overline{M}$ is the matrix formed by conjugating each of the entries of M. The transpose of $\overline{M}$ will be denoted by M^H, i.e. $M^H = (\overline{M})^T$. The vector space of all $m \times n$ matrices with complex entries is denoted by $\mathbb{C}^{m\times n}$.

Theorem 6.4.4. *If A and B are elements of $\mathbb{C}^{m\times n}$, then the following properties are easily verified:*

1. $(A^H)^H = A$
2. $(\alpha A + \beta B)^H = \overline{\alpha}A^H + \overline{\beta}B^H$
3. $(AC)^H = C^H A^H$.

Definition 6.4.5 (Hermitian Matrix)**.** *A matrix M is said to be Hermitian if $M = M^H$.*

Example 6.4.6. The matrix

$$M = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix}$$

is Hermitian since

$$M^H = \begin{bmatrix} \overline{3} & \overline{2-i} \\ \overline{2+i} & \overline{4} \end{bmatrix}^T = \begin{bmatrix} 3 & 2+i \\ 2-i & 4 \end{bmatrix}^T = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix} = M$$

ↂ

If M is a matrix with real entries, then $M^H = M^T$. In particular, if M is a real symmetric matrix, then M is Hermitian. Thus we may view Hermitian matrices as the complex analogue of real symmetric matrices. Hermitian matrices have many nice properties, as we shall see in the next theorem.

Theorem 6.4.7. *The eigenvalues of a Hermitian matrix are all real. Furthermore, eigenvectors belonging to distinct eigenvalues are orthogonal.*

Proof. Let A be a Hermitian matrix. Let λ be an eigenvalue of A and let $\mathbf{x}$ be an eigenvector belonging to λ. If $\alpha = \mathbf{x}^H A\mathbf{x}$, then

$$\overline{\alpha} = \alpha^H = (\mathbf{x}^H A\mathbf{x})^H = \mathbf{x}^H A\mathbf{x} = \alpha$$

Thus, α is real. It follows that

$$\alpha = \mathbf{x}^H A\mathbf{x} = \mathbf{x}^H \lambda\mathbf{x} = \lambda\|\mathbf{x}\|^2$$

and hence

$$\lambda = \frac{\alpha}{\|\mathbf{x}\|^2}$$

is real. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are eigenvectors belonging to distinct eigenvalues λ_1 and λ_2, respectively, then

$$(A\mathbf{x}_1)^H\mathbf{x}_2 = \mathbf{x}_1^H A^H\mathbf{x}_2 = \mathbf{x}_1^H A\mathbf{x}_2 = \lambda_2\mathbf{x}_1^H\mathbf{x}_2$$

On the other hand,

$$(A\mathbf{x}_1)^H\mathbf{x}_2 = (\lambda_1\mathbf{x}_1)^H\mathbf{x}_2 = \overline{\lambda}_1\mathbf{x}_1^H\mathbf{x}_2 = \lambda_1\mathbf{x}_1^H\mathbf{x}_2$$

since the eigenvalues of A are all real. The difference of two equations give

$$(\lambda_1 - \lambda_2)\mathbf{x}_1^H\mathbf{x}_2 = 0 \quad \rightarrow \quad \mathbf{x}_1^H\mathbf{x}_2 = \langle \mathbf{x}_2, \mathbf{x}_1\rangle = 0$$

due to the fact that $\lambda_1 \neq \lambda_2$.

□

Definition 6.4.8 (Unitary Matrix)**.** *An $n \times n$ matrix U is said to be unitary if its column vectors form an orthonormal set in $\mathbb{C}^n$.*

Theorem 6.4.9. *An $n \times n$ matrix U is unitary if and only if $U^H U = I$.*

Remark 6.4.10. *A real unitary matrix is an orthogonal matrix.*

Corollary 6.4.11. *If the eigenvalues of a Hermitian matrix A are distinct, then there exists a unitary matrix U that diagonalizes A.*

Proof. Corollary 6.3.6 + Thm 6.4.7. Construct a diagonal matrix D and a unitary matrix U where the k-th diagonal entry of D is an eigenvalue of A corresponding to the eigenvector, the k-th column vector of the unitary matrix U such that $A = UDU^H$. □

Example 6.4.12. Let

$$A = \begin{bmatrix} 2 & 1-i \\ 1+i & 1 \end{bmatrix}$$

Find a unitary matrix U that diagonalizes A.

The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = 0$, with corresponding eigenvectors $\mathbf{x}_1 = (1-i, 1)^T$ and $\mathbf{x}_2 = (-1, 1+i)^T$. Let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1-i \\ 1 \end{pmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1+i \end{pmatrix}$$

Thus

$$U = \frac{1}{\sqrt{3}}\begin{bmatrix} 1-i & -1 \\ 1 & 1+i \end{bmatrix}$$

$D = U^H AU = diag(3, 0)$ and

$$A = UDU^H = \frac{1}{3}\begin{bmatrix} 1-i & -1 \\ 1 & 1+i \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1+i & 1 \\ -1 & 1-i \end{bmatrix} = \begin{bmatrix} 2 & 1-i \\ 1+i & 1 \end{bmatrix}.$$

ꝏ

Theorem 6.4.13 (Schur's Theorem)**.** *For each $n \times n$ matrix A, there exists a unitary matrix U such that $U^H AU$ is upper triangular.*

The factorization $A = UTU^H$ is often referred to as the *Schur decomposition* of A. In the case that A is Hermitian, the matrix T will be diagonal.

Theorem 6.4.14 (Spectral Theorem)**.** *If A is Hermitian, then there exists a unitary matrix U that diagonalizes A. Namely, $A = UDU^H$*

Proof. By Theorem 6.4.13, there is a unitary matrix U such that $U^H AU = T$, where T is upper triangular. Furthermore,

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

Therefore, T is Hermitian and consequently must be diagonal. □

Example 6.4.15. Let

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

Find the Schur's decomposition of A and B.

The matrices A and B both have the same eigenvalues

$$\lambda_1 = 4, \qquad \lambda_2 = \lambda_3 = 2$$

The eigenspace of A corresponding to $\lambda_1 = 4$ is spanned by $\mathbf{e}_2$, and the eigenspace corresponding to $\lambda = 2$ is spanned by $\mathbf{e}_3$. Since A has only two linearly independent eigenvectors, it is defective. Notice that $\mathbf{e}_2$ and $\mathbf{e}_3$ are already orthogonal unit vectors. We enlarge the subspace spanned by $\mathbf{e}_2$ and $\mathbf{e}_3$ to $\mathbb{C}^3$ by adding a third linearly independent vector $\mathbf{x}_3 = \mathbf{e}_1$. Since $\{\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1\}$ form an orthonormal basis for $\mathbb{C}^3$, then define

$$U_A = [\mathbf{e}_2|\mathbf{e}_3|\mathbf{e}_1], \quad T_A = U_A^H A U_A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus, $A = U_A T U_A^H$ gives one possible Schur decomposition. The unitary matrix can also be formed in the following way

$$U_A = [\mathbf{e}_3|\mathbf{e}_2|\mathbf{e}_1], \quad T_A = U_A^H A U_A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

In either way, we notice that T_A is upper triangular and we always put

On the other hand, the matrix B has eigenvector $\mathbf{x}_1 = (0, 2, 1)^T$ corresponding to $\lambda_1 = 4$ and eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_3$ corresponding to $\lambda = 2$. Thus, B has three linearly independent eigenvectors and consequently is not defective.

Perform Gram-Schmidt Process on the set $\{\mathbf{e}_1, \mathbf{e}_3, \mathbf{x}_1\}$ Let $\mathbf{u}_1 = \mathbf{e}_1$ and $\mathbf{u}_2 = \mathbf{e}_3$, then by take

$$\mathbf{v}_3 = \mathbf{x}_1 - (\mathbf{x}_1^T \mathbf{u}_2)\mathbf{u}_2 - (\mathbf{x}_1^T \mathbf{u}_1)\mathbf{u}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{u}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$U_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad T_B = U_B^H B U_B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

$$B = U_B T_B U_B^H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

Theorem 6.4.16 (Real Schur's Theorem**). *If A is an $n \times n$ matrix with real entries, then A can be factored into a product QTQ^T , where Q is an orthogonal matrix and T is in Schur form:*

$$T = \begin{bmatrix} B_1 & * & \cdots & * \\ & B_2 & & * \\ & O & \ddots & \\ & & & B_j \end{bmatrix}$$

where the B_is are either 1×1 or 2×2 matrices. Each 2×2 block will correspond to a pair of complex conjugate eigenvalues of A. The matrix T is referred to as the real Schur form of A.

Definition 6.4.17. *A matrix A is said to be **orthogonally diagonalizable** if there exists an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^T$.*

Theorem 6.4.18 (Real Spectral Theorem). *A matrix A is orthogonally diagonalizable if and only if it is real symmetric; that is, $Q^T AQ = D$, where D is diagonal.*

Example 6.4.19. Let

$$C = \begin{bmatrix} 0 & 2 & -1 \\ 2 & 3 & -2 \\ -1 & -2 & 0 \end{bmatrix}$$

Find an orthogonal matrix Q that diagonalizes C.

The matrices C has eigenvalues $\lambda_1 = 5$ and $\lambda_2 = \lambda_3 = -1$ since

$$\det(C) = \begin{vmatrix} -\lambda & 2 & -1 \\ 2 & 3-\lambda & -2 \\ -1 & -2 & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda-1 & 2 & -1 \\ 0 & 3-\lambda & -2 \\ -1-\lambda & -2 & -\lambda \end{vmatrix} = -(1+\lambda)\begin{vmatrix} 1 & 2 & -1 \\ 0 & 3-\lambda & -2 \\ 1 & -2 & -\lambda \end{vmatrix}$$

$$= -(1+\lambda)\begin{vmatrix} 1 & 2 & -1 \\ 0 & 3-\lambda & -2 \\ 0 & -4 & -\lambda+1 \end{vmatrix} = -(1+\lambda)\begin{vmatrix} 3-\lambda & -2 \\ -4 & -\lambda+1 \end{vmatrix} = -(1+\lambda)(\lambda^2-4\lambda-5)$$

The eigenspace of C corresponding to $\lambda_1 = 5$ is spanned by $\mathbf{x}_1 = (-1,-2,1)^T$, and the eigenspace corresponding to $\lambda = -1$ is spanned by $\mathbf{x}_2 = (1,0,1)^T$ and $\mathbf{x}_3 = (-2,1,0)^T$. To form an orthogonal matrix, we use Gram-Schmidt process to find an orthonormal basis for both eigenspaces. Let

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

and

$$\mathbf{v}_3 = \mathbf{x}_3 - (\mathbf{x}_3^T\mathbf{u}_2)\mathbf{u}_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

Since $\mathbf{x}_1$ must be orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$ (Theorem 6.4.7), we need only normalize $\mathbf{x}_3$.

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

$$Q = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{3} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}, \quad D = Q^TCQ = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

ↀ

To summarize all the good properties about a real symmetric matrix, we present the following theorem

Theorem 6.4.20. *If A is a real symmetric matrix, then*

1. *the eigenvalues of A are real,*
2. *A must be orthogonally diagonalizable,*
3. *the eigenvectors corresponding to distinct eigenvalues are orthogonal.*

6.5 Singular Value Decomposition

If A is a rectangular matrix, then it has no eigenvalues. Can we still "diagonalize" it? We will present a method for determining how close A is to a matrix of smaller rank.

Theorem 6.5.1. *For any $m \times n$ matrix A of rank r, we have the **singular value decomposition** (SVD):*

$$A = U\Sigma V^T, \tag{6.1}$$

where

1. *U is an $m \times m$ orthogonal matrix,*
2. *V is an $n \times n$ orthogonal matrix, and*
3. *Σ is an $m \times n$ matrix given by*

$$\Sigma = \begin{bmatrix} D & \mathbf{O} \end{bmatrix}, \text{ or } \Sigma = \begin{bmatrix} D \\ \mathbf{O} \end{bmatrix},$$

where $D = diag\{\sigma_1, \ldots, \sigma_k\}$, $k = \min\{m, n\}$ and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_k = 0.$$

*where the σ_i's determined by this factorization are unique and are called the **singular values** of A.*

We will show that the rank of A equals the number of nonzero singular values, and that the magnitudes of the nonzero singular values provide a measure of how close A is to a matrix of lower rank.

Suppose A is an $m \times n$ matrix with rank $r < \min\{m, n\}$. The matrix AA^T will be $m \times m$ with rank r. The matrix A^TA will be $n \times n$ with rank r. Both matrices A^TA and AA^T will be *positive semidefinite* (See in Sec 6.7), and will therefore have r (possibly repeated) positive eigenvalues, and r linearly independent corresponding eigenvectors. If we set

$$U_1 = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r], \qquad V = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r]$$

and define $\Sigma_1 = \operatorname{diag}\{\sigma_1, \ldots, \sigma_r\}$ with positive diagonal entries.Then

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & O \\ O & O \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T \tag{6.2}$$

The factorization (6.2) is called the *compact form of the singular value decomposition of A*. This form is useful in many applications, such as digital image compression, principal component analysis and etc. This singular value decomposition has a particularly nice representation if we carry through the multiplication of the matrices as

$$A = U_1 \Sigma V_1^T = \mathbf{u}_1 \sigma_1 \mathbf{v}_1^T + \cdots + \mathbf{u}_r \sigma_r \mathbf{v}_r^T$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} column of U_1 and V_1 corresponding to σ_i, respectively. Then each of these pieces has rank 1. If we order the singular values

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

then the singular value decomposition gives A in r rank 1 pieces in *order of importance*.

We can see the matrices A^TA and AA^T are the key ingredients in forming the singular value decomposition. These two matrices are real symmetric and positive (semi)definite (Sec 6.7). We will start from showing some good properties of them.

Theorem 6.5.2. *Let A be an $m \times n$ matrix, where $m \neq n$, then*

1. *A^TA and AA^T are real symmetric matrices, orthogonally diagonalizable.*

2. *all the eigenvalues of A^TA and AA^T are non-negative and real.*

3. *rank$(A^TA) =$ rank$(AA^T) =$ rank$(A^T) =$ rank(A).*

4. *the nonzero eigenvalues of A^TA and AA^T are the same.*

   More generally, for any $n \times m$ matrix B and $m \times n$ matrix C, BC and CB share the same nonzero eigenvalues.

Proof. 1 and 3 as exercises.

2. Let λ be an eigenvalue of A^TA with an eigenvector $\mathbf{x}$, then

$$\mathbf{x}^T(A^TA\mathbf{x}) = \mathbf{x}^T(\lambda\mathbf{x}) = \lambda\mathbf{x}^T\mathbf{x} \quad \to \quad \lambda = \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \geq 0$$

   since $\mathbf{x} \neq \mathbf{0}$ and $\|A\mathbf{x}\| \geq 0$.

   A similar proof will work for AA^T.

4. Assume $\lambda \neq 0$ is an eigenvalue of A^TA with an eigenvector $\mathbf{x}$, then

$$A^TA\mathbf{x} = \lambda\mathbf{x} \quad \to \quad AA^T(A\mathbf{x}) = A\lambda\mathbf{x} \quad \to \quad AA^T(A\mathbf{x}) = \lambda(A\mathbf{x})$$

   Recall that $\mathrm{N}(A^TA) = \mathrm{N}(A)$, if $\mathbf{x} \in \mathrm{N}(A)$, then $\lambda\mathbf{x} = A^TA\mathbf{x} = \mathbf{0}$, impossible. So $\mathbf{x} \notin \mathrm{N}(A)$ and $A\mathbf{x} \neq \mathbf{0}$. Hence, λ is an eigenvalue of AA^T with an eigenvector $A\mathbf{x}$.

□

Theorem 6.5.3. *The rank of A equals the number of nonzero singular values of A.*

Proof. The matrix AA^T is real symmetric, then it is diagonalized by some orthogonal matrix Q, i.e. $AA^T = QDQ^T$.

$$\mathrm{rank}(A) = \mathrm{rank}(AA^T) = \mathrm{rank}(QDQ^T) = \mathrm{rank}(D)$$

$=$ the # of nonzero eigenvalues of AA^T $=$ the # of nonzero singular values of A.

□

Theorem 6.5.4. *If matrix A of rank r has a singular value decomposition, $U\Sigma V^T$. Then*

1. *The first r columns of V form an orthonormal basis for $Col\big(A^T\big)$.*
2. *The rest of columns in V form an orthonormal basis for $N(A)$.*
3. *The first r columns of U form an orthonormal basis for $Col(A)$.*
4. *The rest of columns in U form an orthonormal basis for $N\big(A^T\big)$.*

Proof. 1-2 as exercises.

3. Let $\lambda_i \neq 0$ be an eigenvalue of AA^T with an eigenvector $\mathbf{u}_i$, for $i = 1, \cdots, r$, then

$$AA^T\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad \rightarrow \quad \mathbf{u}_i = \frac{1}{\lambda_i}A(A^T\mathbf{u}_i) \in \mathrm{Col}(A)$$

since $A(A^T\mathbf{u}_i) \in \mathrm{Col}(A)$.

4. Assume $\mathbf{u}_j$ is an eigenvector of A^TA corresponding to zero eigenvalues for $j = r+1, \cdots, m$, then

$$AA^T\mathbf{u}_j = \mathbf{0} \quad \rightarrow \quad \mathbf{u}_j \in \mathrm{N}(AA^T) = \mathrm{N}(A^T)$$

□

Steps on finding a singular value decomposition

Step 1 Form Σ with singular values of A: $\sigma_i = \sqrt{\text{eigenvalues of } A^TA}$

Step 2 Columns in V: eigenvectors of A^TA

Step 3 Columns in U: $\mathbf{u}_i = \begin{cases} \frac{1}{\sigma_i}A\mathbf{v}_i & i = 1, \cdots, r \\ \in \mathrm{N}(A^T) & i = r+1, \cdots, m \end{cases}$

Example 6.5.5. Find an SVD for $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$.

The matrix

$$A^T A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 0$. Consequently, the singular values of A are $\sigma_1 = \sqrt{4} = 2$ and $\sigma_2 = 0$. The eigenvalue λ_1 has eigenvectors of the form $\alpha(1,1)^T$, and λ_2 has eigenvectors of the form $\beta(1,-1)^T$. Therefore, the orthogonal matrix

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

diagonalizes $A^T A$. By observation, it follows that

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

The remaining column vectors of U must form an orthonormal basis for $\mathrm{N}(A^T)$. We can compute a basis $\{\mathbf{x}_2, \mathbf{x}_3\}$ for $\mathrm{N}(A^T)$ in the usual way.

$$\mathbf{x}_2 = (1,-1,0)^T, \quad \mathbf{x}_3 = (0,0,1)^T$$

Since these vectors are already orthogonal, it is not necessary to use the GramSchmidt process to obtain an orthonormal basis. We need only set

$$\mathbf{u}_2 = \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|} = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)^T, \qquad \mathbf{u}_3 = \mathbf{x}_3 = (0,0,1)^T$$

It then follows that

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

ꝏ

Exercise 6.5.6.

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

✓

The matrix U and V are not unique.

6.7 Positive Definite Matrices

Definition 6.7.1 (Definite Matrices)**.** *An $n \times n$ real symmetric matrix A is said to be*

1. ***positive definite*** *if $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$,*
2. ***negative definite*** *if $\mathbf{x}^T A\mathbf{x} < 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$,*
3. ***positive semidefinite*** *if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$,*
4. ***negative semidefinite*** *if $\mathbf{x}^T A\mathbf{x} \leq 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$,*
5. ***indefinite*** *if $\mathbf{x}^T A\mathbf{x}$ takes on values that differ in sign.*

Properties of Positive Definite Matrices

Theorem 6.7.2 (Eigenvalue Test)**.** *A matrix A is positive definite if and only if all its eigenvalues are positive.*

Theorem 6.7.3. *If A is positive definite, then A is nonsingular, and* $\det A > 0$.

Proof. Prove as examples □

Definition 6.7.4 (Leading Principal Submatrix)**.** *Given an $n \times n$ matrix A, the* **leading principal submatrix** A_r *of A of order r is the matrix formed by deleting the last $n-r$ rows and columns of A.*

Theorem 6.7.5 (Determinant Test)**.** *A symmetric matrix A is positive definite if and only if every leading principal submatrice of A has positive determinant.*

Theorem 6.7.6. *Let A be a symmetric $n \times n$ matrix. The following are equivalent:*

(a) A is positive definite.

(b) The leading principal submatrices $A_1, \cdots, A_n$ all have positive determinants.

*(c) ** A can be reduced to upper triangular form with positive diagonal entries using only row operation III.*

*(d) ** (Cholesky factorization) $A = LL^T$ where L is lower triangular with positive diagonal entries.*

*(e) ** $A = B^T B$ for some nonsingular matrix B.*

6.6 Quadratic Forms

Definition 6.6.1. *A quadratic equation of two variables x, y is an equation of the form*

$$ax^2 + by^2 + cxy + dx + ey + f = 0, \tag{6.3}$$

where a, b, c, d, e and f are constants. The term

$$\mathbf{x}^T A\mathbf{x} = ax^2 + by^2 + cxy$$

*is called a **quadratic form** of two variables associated with* (6.3)*.*

Remark 6.6.2 (3-Dimensional Case)**.** *A quadratic equation of three variables x, y, z is an equation of the form*

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + mz + n = 0 \tag{6.4}$$

for some $a, b, c, d, e, f, g, h, m, n$. The term

$$Q(x, y, z) = \mathbf{x}^T A\mathbf{x} = ax^2 + by^2 + cz^2 + dxy + exz + fyz$$

is called the ***quadratic form*** *of the three variables associated with* (6.4)*.*

Definition 6.6.3 (n-Dimensional Case)**.** *A quadratic form in n variables can always be denoted as*

$$Q(\mathbf{x}) = \mathbf{x^T A x}$$

for some symmetric matrix A.

Definition 6.6.4 (Quadratic equation)**.** *A quadratic equation in n variables $x_1, \cdots, x_n$ is one of the form $\mathbf{x}^T A\mathbf{x} + B\mathbf{x} + \alpha = 0$ where $\mathbf{x} = (x_1, \cdots, x_n)^T$, A is an $n \times n$ symmetric matrix, B is a $1 \times n$ matrix, and α is a scalar.*

Remark 6.6.5. *In a two-dimensional case, if A is nonsingular, then by rotating and translating the axes, it is possible to rewrite the equation into the standard form*

$$\lambda_1(x_{new})^2 + \lambda_2(y_{new})^2 + \alpha_{new} = 0,$$

where λ_1 and λ_2 are the eigenvalues of A.

For the general n-dimensional case, the quadratic form can always be translated to a simpler diagonal form. More precisely, we have the following theorem:

Theorem 6.6.6 (Principal Axes Theorem)**.** *If A is a real symmetric matrix, then there is a change of variables $\mathbf{y} = Q^T\mathbf{x}$ so that $\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y}$, where D is a diagonal matrix.*

Optimization: An Application to Calculus

Hessian matrix and local extrema

Theorem 6.6.7. *If A is nonsingular, then $\mathbf{x}_0$ is the only critical point of the function $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, and*

1. *$\mathbf{x}_0$ is the global minimum of Q if $H(\mathbf{x}_0)$ is positive definite.*

2. *$\mathbf{x}_0$ is the global maximum of Q if $H(\mathbf{x}_0)$ is negative definite.*

3. *$\mathbf{x}_0$ is a saddle point of Q if $H(\mathbf{x}_0)$ is indefinite.*