8. Diagonalization
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

We know that every linear transformation \( T: \mathbb{R}^n \rightarrow \mathbb{R}^m \) has an associated standard matrix

\[
[T] = [T(e_1) \mid T(e_2) \mid \cdots \mid T(e_n)]
\]

with the property that

\[
T(x) = [T]x
\]

for every vector \( x \) in \( \mathbb{R}^n \). For the moment we will focus on the case where \( T \) is a linear operator on \( \mathbb{R}^n \), so the standard matrix \([T]\) is a square matrix of size \( n \times n \).

Sometimes the form of the standard matrix fully reveals the geometric properties of a linear operator and sometimes it does not.

\[
[T_1] = \begin{bmatrix}
\cos \frac{\pi}{4} & -\sin \frac{\pi}{4} & 0 \\
\sin \frac{\pi}{4} & \cos \frac{\pi}{4} & 0 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
0 & 0 & 1
\end{bmatrix} \quad [T_2] = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
\]

Our primary goal in this section is to develop a way of using bases other than the standard basis to create matrices that describe the geometric behavior of a linear transformation better than the standard matrix.
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

Suppose that
\[ \mathbf{x} \xrightarrow{T} T(\mathbf{x}) \]
is a linear operator on \( \mathbb{R}^n \) and \( B \) is a basis for \( \mathbb{R}^n \). In the course of mapping \( \mathbf{x} \) into \( T(\mathbf{x}) \) this operator creates a companion operator
\[
[\mathbf{x}]_B \xrightarrow{\cdot} [T(\mathbf{x})]_B
\]
that maps the coordinate matrix \([\mathbf{x}]_B\) into the coordinate matrix \([T(\mathbf{x})]_B\).

There must be a matrix \( A \) such that
\[
A[\mathbf{x}]_B = [T(\mathbf{x})]_B
\]
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

**Theorem 8.1.1** Let \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) be a linear operator, let \( B = \{ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \} \) be a basis for \( \mathbb{R}^n \), and let

\[
A = \begin{bmatrix}
[T(\mathbf{v}_1)]_B & [T(\mathbf{v}_2)]_B & \cdots & [T(\mathbf{v}_n)]_B
\end{bmatrix}
\]  

(4)

Then

\[
[T(\mathbf{x})]_B = A[\mathbf{x}]_B
\]

(5)

for every vector \( \mathbf{x} \) in \( \mathbb{R}^n \). Moreover, the matrix \( A \) given by Formula (4) is the only matrix with property (5).

Let \( \mathbf{x} \) be any vector in \( \mathbb{R}^n \), and suppose that its coordinate matrix with respect to \( B \) is

\[
[\mathbf{x}]_B = \begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
c_n
\end{bmatrix}
\]

That is, \( \mathbf{x} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n \).
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

It now follows from the linearity of $T$ that

$$T(x) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n)$$

and from the linearity of coordinate maps that

$$[T(x)]_B = c_1 [T(v_1)]_B + c_2 [T(v_2)]_B + \cdots + c_n [T(v_n)]_B$$

which we can write in matrix form as

$$[T(x)]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = A[x]_B$$

This proves that the matrix $A$ in (4) has property (5). Moreover, $A$ is the only matrix with this property, for if there exists a matrix $C$ such that

$$[T(x)]_B = A[x]_B = C[x]_B$$

for all $x$ in $\mathbb{R}^n$, then Theorem 7.11.6 implies that $A=C$. 
8.1. Matrix Representations of Linear Transformations

Matrix of a Linear Operator with Respect to a Basis

The matrix $A$ in (4) is called the \textit{matrix for $T$ with respect to the basis $B$} and is denoted by

$$[T]_B = \begin{bmatrix} [T(v_1)]_B & [T(v_2)]_B & \cdots & [T(v_n)]_B \end{bmatrix}$$

Using this notation we can write (5) as

$$[T(x)]_B = [T]_B [x]_B$$
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

Example 1

Let $T: R^2 \to R^2$ be the linear operator whose standard matrix is

$$[T] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

Find the matrix for $T$ with respect to the basis $B=\{v_1,v_2\}$, where

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

The images of the basis vectors under the operator $T$ are

$$T(v_1) = [T]v_1 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = v_1 = v_1 + 0v_2$$

$$T(v_2) = [T]v_2 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} \end{bmatrix} = 5v_2 = 0v_1 + 5v_2$$
Matrix of A Linear Operator with Respect to A Basis

Example 1

so the coordinate matrices of these vectors with respect to $B$ are

$$[T(v_1)]_B = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad [T(v_2)]_B = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

Thus, it follows from (6) that

$$[T]_B = \begin{bmatrix} [T(v_1)]_B \mid [T(v_2)]_B \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

This matrix reveals geometric information about the operator $T$ that was not evident from the standard matrix. It tells us that the effect of $T$ is to stretch the $v_2$-coordinate of a vector by a factor of 5 and to leave the $v_1$-coordinate unchanged.
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Operator with Respect to A Basis

Example 2

Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the linear operator whose standard matrix is

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

We showed in Example 7 of Section 6.2 that $T$ is a rotation through an angle $2\pi/3$ about an axis in the direction of the vector $\mathbf{n} = (1,1,1)$.

Let us now consider how the matrix for $T$ would look with respect to an orthonormal basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in which $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$ is a positive scalar multiple of $\mathbf{n}$ and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthonormal basis for the plane $W$ through the origin that is perpendicular to the axis of rotation.
Matrix of A Linear Operator with Respect to A Basis

Example 2

The rotation leaves the vector $v_3$ fixed, so

$$T(v_3) = v_3 = 0v_1 + 0v_2 + 1v_3$$

and hence

$$[T(v_3)]_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Also, $T(v_1)$ and $T(v_2)$ are linear combinations of $v_1$ and $v_2$, since these vectors lie in $W$. This implies that the third coordinate of both $[T(v_1)]_B$ and $[T(v_2)]_B$ must be zero, and the matrix for $T$ with respect to the basis $B$ must be of the form

$$[T]_B = 
\begin{bmatrix}
[T(v_1)]_B & [T(v_2)]_B & [T(v_3)]_B
\end{bmatrix}
= \begin{bmatrix}
\times & \times & 0 \\
\times & \times & 0 \\
0 & 0 & 1
\end{bmatrix}$$
Matrix of A Linear Operator with Respect to A Basis

Example 2

Since \( T \) behaves exactly like a rotation of \( R^2 \) in the plane \( W \), the block of missing entries has the form of a rotation matrix in \( R^2 \). Thus,

\[
[T]_B = \begin{bmatrix} [T(v_1)]_B & [T(v_2)]_B & [T(v_3)]_B \end{bmatrix} = \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} & 0 \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]
8.1. Matrix Representations of Linear Transformations

Changing Bases

Suppose that $T: R^n \rightarrow R^n$ is a linear operator and that $B = \{v_1, v_2, \ldots, v_n\}$ and $B' = \{v'_1, v'_2, \ldots, v'_n\}$ are bases for $R^n$. Also, let $P_{B \rightarrow B'}$ be the transition matrix from $B$ to $B'$ (so $P_{B' \rightarrow B}^{-1}$ is the transition matrix from $B'$ to $B$).

Changing Bases

![Diagram](image)

The diagram shows two different paths from $[x]_B$ to $[T(x)]_B$:

1. $[T]_{B'} [x]_{B'} = [T(x)]_{B'}$ (10)
   
   (i) Multiply $[x]_{B'}$ on the left by $P_{B \rightarrow B'}^{-1}$ to obtain $P_{B \rightarrow B'}^{-1} [x]_{B'} = [x]_B$.

2. Multiply $[x]_B$ on the left by $[T]_B$ to obtain $[T]_B [x]_B = [T(x)]_B$. (11)
   
   (ii) Multiply $[T(x)]_B$ on the left by $P$ to obtain $[T(x)]_B [x]_B = [T(x)]_B$.

Thus, (10) and (11) together imply that $(P [T]_B P_{B' \rightarrow B}^{-1}) [x]_{B'} = [T]_{B'} [x]_{B'}$.

Since this holds for all $x$ in $R^n$, it follows from Theorem 7.11.6 that $P [T]_B P_{B' \rightarrow B}^{-1} = [T]_{B'}$. 

E-mail: hogijung@hanyang.ac.kr

http://web.yonsei.ac.kr/hgjung
8.1. Matrix Representations of Linear Transformations

Changing Bases

**Theorem 8.1.2** If $T : R^n \to R^n$ is a linear operator, and if $B = \{v_1, v_2, \ldots, v_n\}$ and $B' = \{v'_1, v'_2, \ldots, v'_n\}$ are bases for $R^n$, then $[T]_B$ and $[T]_{B'}$ are related by the equation

$$[T]_{B'} = P[T]_B P^{-1}$$

(12)

in which

$$P = P_{B \to B'} = \begin{bmatrix} [v_1]_{B'} & [v_2]_{B'} & \cdots & [v_n]_{B'} \end{bmatrix}$$

(13)

is the transition matrix from $B$ to $B'$. In the special case where $B$ and $B'$ are orthonormal bases the matrix $P$ is orthogonal, so (12) is of the form

$$[T]_{B'} = P[T]_B P^T$$

(14)

When convenient, Formula (12) can be rewritten as

$$[T]_B = P^{-1} [T]_{B'} P$$

and in the case where the bases are orthonormal this equation can be expressed as

$$[T]_B = P^T [T]_{B'} P$$
8.1. Matrix Representations of Linear Transformations

Changing Bases

Since many linear operators are defined by their standard matrices, it is important to consider the special case of Theorem 8.1.2 in which $B' = S$ is the standard basis for $\mathbb{R}^n$.

In this case $[T]_B = [T]_S = [T]$, and the transition matrix $P$ from $B$ to $B'$ has the simplified form

$$ P = P_{B \to B'} = P_{B \to S} = \begin{bmatrix} [v_1]_S & [v_2]_S & \cdots & [v_n]_S \end{bmatrix} = [v_1 \mid v_2 \mid \cdots \mid v_n] $$
8.1. Matrix Representations of Linear Transformations

Changing Bases

**Theorem 8.1.3** If \( T : \mathbb{R}^n \to \mathbb{R}^n \) is a linear operator, and if \( B = \{v_1, v_2, \ldots, v_n\} \) is a basis for \( \mathbb{R}^n \), then \([T]\) and \([T]_B\) are related by the equation

\[
[T] = P[T]_B P^{-1}
\]

in which

\[
P = [v_1 \mid v_2 \mid \cdots \mid v_n]
\]

is the transition matrix from \( B \) to the standard basis. In the special case where \( B \) is an orthonormal basis the matrix \( P \) is orthogonal, so (17) is of the form

\[
[T] = P[T]_B P^T
\]

When convenient, Formula (17) can be rewritten as

\[
[T]_B = P^{-1}[T] P
\]

and in the case where \( B \) is an orthonormal basis this equation can be expressed as

\[
[T]_B = P^T[T] P
\]
8.1. Matrix Representations of Linear Transformations

Changing Bases

Formula (17) [or (19) in the orthogonal case] tells us that the process of changing from the standard basis for $\mathbb{R}^n$ to a basis $B$ produces a factorization of the standard matrix for $T$ as

$$[T] = P[T]_B P^{-1}$$

in which $P$ is the transition matrix from the basis $B$ to the standard basis $S$. To understand the geometric significance of this factorization, let us use it to compute $T(x)$ by writing

$$T(x) = [T]x = (P[T]_B P^{-1})x = P[T]_B (P^{-1}x)$$

Reading from right to left, this equation tells us that $T(x)$ can be obtained by first mapping the standard coordinates of $x$ to $B$-coordinates using the matrix $P^{-1}$, then performing the operation on the $B$-coordinates using the matrix $[T]_B$, and then using the matrix $P$ to map the resulting vector back to standard coordinates.
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 3

In Example 1 we considered the linear operator \( T: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) whose standard matrix is

\[
A = [T] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}
\]

and we showed that

\[
[T]_B = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}
\]

with respect to the orthonormal basis \( B=\{v_1,v_2\} \) that is formed from the vectors

\[
v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}
\]
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 3

In this case the transition matrix from $B$ to $S$ is

$$P = [v_1 \mid v_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

so it follows from (17) that $[T]$ can be factored as

$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$[T] = P [T]_B P^{-1}$$
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 4

In Example 2 we considered the rotation $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ whose standard matrix is

$$A = [T] = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

and we showed that

$$[T]_B = \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} & 0 \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

From Example 7 of Section 6.2, the equation of the plane $W$ is $x + y + z = 0$

From Example 10 of Section 7.9, an orthonormal basis for the plane is

$$v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$$
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 4

Since \( \mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2 \),

\[
\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2 = \begin{bmatrix}
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
-\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}}
\end{bmatrix} = \frac{1}{\sqrt{3}} \mathbf{i} + \frac{1}{\sqrt{3}} \mathbf{j} + \frac{1}{\sqrt{3}} \mathbf{k}
\]

The transition matrix from \( B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \) to the standard basis is

\[
P = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3] = \begin{bmatrix}
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{bmatrix}
\]

Since this matrix is orthogonal, it follows from (19) that \([T]\) can be factored as

\[
\begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix} = \begin{bmatrix}
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{bmatrix} \begin{bmatrix}
\cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} & 0 \\
\sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix} \begin{bmatrix}
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{bmatrix} P^T
\]

\[
[T] = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix} \begin{bmatrix}
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{bmatrix} P^T
\]

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 5

From Formula (2) of Section 6.2, the standard matrix for the reflection $T$ of $R^2$ about the line $L$ through the origin making an angle $\theta$ with the positive $x$-axis of a rectangular $xy$-coordinate system is

$$[T] = H_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$$

Suppose that we rotate the coordinate axes through the angle $\theta$ to align the new $x'$-axis with $L$, and we let $v_1$ and $v_2$ be unit vectors along the $x'$- and $y'$-axes, respectively.

Since

$$T(v_1) = v_1 = v_1 + 0v_2 \quad \text{and} \quad T(v_2) = -v_2 = 0v_1 + (-1)v_2$$

it follows that the matrix for $T$ with respect to the basis $B=\{v_1,v_2\}$ is

$$[T]_B = \begin{bmatrix} [T(v_1)]_B | [T(v_2)]_B \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
8.1. Matrix Representations of Linear Transformations

Changing Bases

Example 5

Also, it follows from Example 8 of Section 7.11 that the transition matrices between the standard basis $S$ and the basis $B$ are

\[
P = P_{B \to S} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad P^T = P_{S \to B} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}
\]

Thus, Formula (19) implies that

\[
\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}
\]

\[
[T] = P [T]_B P^T
\]
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Transformation with Respect to A Pair of Bases

Recall that every linear transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ has an associated $m \times n$ standard matrix

$[T] = [T(e_1) \mid T(e_2) \mid \cdots \mid T(e_n)]$

with the property that

$T(x) = [T]x$

If $B$ and $B'$ are bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, then the transformation

$x \xrightarrow{T} T(x)$

creates an associated transformation

$[x]_B \rightarrow [T(x)]_{B'}$

that maps the coordinate matrix $[x]_B$ into the coordinate matrix $[T(x)]_{B'}$. As in the operator case, this associated transformation is linear and hence must be a matrix transformation; that is, there must be a matrix $A$ such that

$A[x]_B = [T(x)]_{B'}$
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Transformation with Respect to A Pair of Bases

**Theorem 8.1.4** Let \( T : R^n \rightarrow R^m \) be a linear transformation, let \( B = \{v_1, v_2, \ldots, v_n\} \) and \( B' = \{u_1, u_2, \ldots, u_m\} \) be bases for \( R^n \) and \( R^m \), respectively, and let

\[
A = \left[ [T(v_1)]_{B'} \mid [T(v_2)]_{B'} \mid \cdots \mid [T(v_n)]_{B'} \right]
\]  
(23)

Then

\[
[T(x)]_{B'} = A[x]_B
\]

(24)

for every vector \( x \) in \( R^n \). Moreover, the matrix \( A \) given by Formula (23) is the only matrix with property (24).

The matrix \( A \) in (23) is called the **matrix for \( T \) with respect to the bases \( B \) and \( B' \)** and is denoted by the symbol \([T]_{B',B'}\).

\[
[T]_{B',B} = \left[ [T(v_1)]_{B'} \mid [T(v_2)]_{B'} \mid \cdots \mid [T(v_n)]_{B'} \right]
\]

\[
[T(x)]_{B'} = [T]_{B',B}[x]_B
\]
8.1. Matrix Representations of Linear Transformations

Matrix of a Linear Transformation with Respect to a Pair of Bases

Example 6

Let \( T: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) be the linear transformation defined by

\[
T \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} x_2 \\ -5x_1 + 13x_2 \\ -7x_1 + 16x_2 \end{bmatrix}
\]

Find the matrix for \( T \) with respect to the bases \( B=\{v_1, v_2\} \) for \( \mathbb{R}^2 \) and \( B'=\{v'_1, v'_2, v'_3\} \) for \( \mathbb{R}^3 \), where

\[
v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}; \quad v'_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad v'_2 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \quad v'_3 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}
\]

Using the given formula for \( T \) we obtain

\[
T(v_1) = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad T(v_2) = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}
\]
8.1. Matrix Representations of Linear Transformations

Matrix of A Linear Transformation with Respect to A Pair of Bases

Example 6

and expressing these vectors as linear combinations of $v'_1$, $v'_2$, and $v'_3$ we obtain

$$T(v_1) = -v'_2 - \frac{3}{2}v'_3$$
$$T(v_2) = \frac{5}{2}v'_1 + \frac{1}{2}v'_2 - \frac{3}{4}v'_3$$

Thus,

$$[T]_{B',B} = 
\begin{bmatrix}
T(v_1)_{B'} & T(v_2)_{B'}
\end{bmatrix} =
\begin{bmatrix}
0 & 5/2 \\
-1 & 1/2 \\
-3/2 & -3/4
\end{bmatrix}$$
8.1. Matrix Representations of Linear Transformations

Effect of Changing Bases on Matrices of Linear Transformations

\[ [T]_{B_1', B_1} = V [T]_{B_2', B_2} U^{-1} \]

In particular, if \( B_1 \) and \( B_1' \) are the standard bases for \( R^n \) and \( R^m \), respectively, and if \( B \) and \( B' \) are any bases for \( R^n \) and \( R^m \), respectively, then it follows from (27) that

\[ [T] = V [T]_{B', B} U^{-1} \]

where \( U \) is the transition matrix from \( B \) to the standard basis for \( R^n \) and \( V \) is the transition matrix from \( B' \) to the standard basis for \( R^m \).
8.1. Matrix Representations of Linear Transformations

Representing Linear Operators with Two Bases

The single-basis representation of $T$ with respect to $B$ can be viewed as the two-basis representation in which both bases are $B$.

$$[T]_B = [T]_{B,B}$$
8.2. Similarity and Diagonalizability

Similar Matrices

**Definition 8.2.1** If $A$ and $C$ are square matrices with the same size, then we say that $C$ is similar to $A$ if there is an invertible matrix $P$ such that $C = P^{-1}AP$.

**REMARK**

If $C$ is similar to $A$, then it is also true that $A$ is similar to $C$. You can see this by letting $Q= P^{-1}$ and rewriting the equation $C= P^{-1}AP$ as

$$A = PCP^{-1} = (P^{-1})^{-1}C(P^{-1}) = Q^{-1}CQ$$
8.2. Similarity and Diagonalizability

Similar Matrices

**Theorem 8.2.2** Two matrices are similar if and only if there exist bases with respect to which the matrices represent the same linear operator.

Let \( T: \mathbb{R}^n \rightarrow \mathbb{R}^n \) denote multiplication by \( A \); that is,

\[
A = [T]
\]  \hspace{1cm} (1)

As \( A \) and \( C \) are similar, there exists an invertible matrix \( P \) such that \( C = P^{-1}AP \), so it follows from (1) that

\[
C = P^{-1}[T]P
\]  \hspace{1cm} (2)

If we assume that the column-vector form of \( P \) is

\[
P = [v_1 \mid v_2 \mid \cdots \mid v_n]
\]

then the invertibility of \( P \) and Theorem 7.4.4 imply that \( B = \{v_1, v_2, \ldots, v_n\} \) is a basis for \( \mathbb{R}^n \). It now follows from Formula (2) above and Formula (20) of Section 8.1 that

\[
C = P^{-1}[T]P = [T]_B
\]
8.2. Similarity and Diagonalizability

**Similarity Invariants**

There are a number of basic properties of matrices that are shared by similar matrices.

For example, if $C = P^{-1}AP$, then

$$\det(C) = \det(P^{-1}AP) = \det(P^{-1}) \det(A) \det(P) = \frac{1}{\det(P)} \det(A) \det(P) = \det(A)$$

which shows that similar matrices have the same determinant.

In general, any property that is shared by similar matrices is said to be a *similarity invariant*. 
8.2. Similarity and Diagonalizability

Similarity Invariants

**Theorem 8.2.3**

(a) Similar matrices have the same determinant.
(b) Similar matrices have the same rank.
(c) Similar matrices have the same nullity.
(d) Similar matrices have the same trace.
(e) Similar matrices have the same characteristic polynomial and hence have the same eigenvalues with the same algebraic multiplicities.

(e) If $A$ and $C$ are similar matrices, then so are $\lambda I - A$ and $\lambda I - C$ for any scalar $\lambda$.

\[
\lambda I - C = \lambda I - P^{-1}AP = \lambda P^{-1}P - P^{-1}AP = P^{-1}(\lambda P - AP) \\
= P^{-1}(\lambda I P - AP) = P^{-1}(\lambda I - A)P
\]

This shows that $\lambda I - A$ and $\lambda I - C$ are similar, so (3) now follows from part (a).

$$\det(\lambda I - C) = \det(\lambda I - A) \quad (3)$$
8.2. Similarity and Diagonalizability

Similarity Invariants

Example 1

Show that there do not exist bases for $\mathbb{R}^2$ with respect to which the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Represent the same linear operator.

For $A$ and $C$ to represent the same linear operator, the two matrices would have to be similar by Theorem 8.2.2. But this cannot be, since $\text{tr}(A)=7$ and $\text{tr}(C)=5$, contradicting the fact that the trace is a similarity invariant.
8.2. Similarity and Diagonalizability

**Eigenvectors and Eigenvalues of Similar Matrices**

Recall that the solution space of

\[(\lambda_0 I - A)x = 0\]

is called the *eigenspace* of \(A\) corresponding to \(\lambda_0\). We call the dimension of this solution space the *geometric multiplicity* of \(\lambda_0\).

Do not confuse this with the *algebraic multiplicity* of \(\lambda_0\), which is the number of repetition of the factor \(\lambda - \lambda_0\) in the complete factorization of the characteristic polynomial of \(A\).
8.2. Similarity and Diagonalizability

Eigenvalues and Eigenvalues of Similar Matrices

Example 2

Find the algebraic and geometric multiplicities of the eigenvalues of

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
1 & 3 & 0 \\
-3 & 5 & 3
\end{bmatrix}
\]

\[
p(\lambda) = \det(\lambda I - A) = \lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)^2
\]

\[\lambda = 2\] has algebraic multiplicity 1

\[\lambda = 3\] has algebraic multiplicity 2

One way to find the geometric multiplicities of the eigenvalues is to find bases for the eigenspaces and then determine the dimensions of those spaces from the number of basis vectors.
8.2. Similarity and Diagonalizability

Eigenvectors and Eigenvalues of Similar Matrices

**Example 2**

$$\begin{bmatrix} \lambda - 2 & 0 & 0 \\ -1 & \lambda - 3 & 0 \\ 3 & -5 & \lambda - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

If $\lambda=2$,

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & -5 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1/8 t \\ -1/8 t \\ t \end{bmatrix} = t \begin{bmatrix} 1/8 \\ -1/8 \\ 1 \end{bmatrix}$$

If $\lambda=3$,

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 3 & -5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ t \end{bmatrix} = t \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$\lambda=2$ has algebraic multiplicity 1, **geometric multiplicity 1**

$\lambda=3$ has algebraic multiplicity 2, **geometric multiplicity 1**
8.2. Similarity and Diagonalizability

Eigenvectors and Eigenvalues of Similar Matrices

Example 3
Find the algebraic and geometric multiplicities of the eigenvalues of

\[ A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \]

\[ p(\lambda) = \det(\lambda I - A) = \lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)^2 \]

\(\lambda = 1\) has algebraic multiplicity 1
\(\lambda = 2\) has algebraic multiplicity 2
8.2. Similarity and Diagonalizability

Eigenvectors and Eigenvalues of Similar Matrices

**Example 3**

\[
\begin{bmatrix}
\lambda & 0 & 2 \\
-1 & \lambda - 2 & -1 \\
-1 & 0 & \lambda - 3 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
0 \\
\end{bmatrix}
\]

If \(\lambda = 1\),
\[
x = \begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix} = 
\begin{bmatrix}
-2t \\
t \\
t \\
\end{bmatrix} = 
\begin{bmatrix}
-2 \\
1 \\
1 \\
\end{bmatrix}
\]

If \(\lambda = 2\),
\[
x = \begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix} = 
\begin{bmatrix}
-s \\
t \\
s \\
\end{bmatrix} = 
\begin{bmatrix}
-s \\
0 \\
0 \\
\end{bmatrix} + 
\begin{bmatrix}
0 \\
t \\
1 \\
\end{bmatrix} = 
\begin{bmatrix}
-1 \\
0 \\
1 \\
\end{bmatrix} + 
\begin{bmatrix}
0 \\
1 \\
0 \\
\end{bmatrix}
\]

\(\lambda = 1\) has algebraic multiplicity 1, geometric multiplicity 1

\(\lambda = 2\) has algebraic multiplicity 2, geometric multiplicity 2
8.2. Similarity and Diagonalizability

Eigenvectors and Eigenvalues of Similar Matrices

**Theorem 8.2.4** Similar matrices have the same eigenvalues and those eigenvalues have the same algebraic and geometric multiplicities for both matrices.

Let us assume first that \( A \) and \( C \) are similar matrices. Since similar matrices have the same characteristics polynomial, it follows that \( A \) and \( C \) have the same eigenvalues with the same algebraic multiplicities.

To show that an eigenvalue has the same geometric multiplicity for both matrices, we must show that the solution spaces of

\[(\lambda I - A)x = 0 \quad \text{and} \quad (\lambda I - C)x = 0\]

have the same dimension, or equivalently, that the matrices

\[\lambda I - A \quad \text{and} \quad \lambda I - C\]

have the same nullity.

But we showed in the proof of Theorem 8.2.3 that the similarity of \( A \) and \( C \) implies the similarity of the matrices in (12). Thus, these matrices have the same nullity by part (c) of Theorem 8.2.3.
8.2. Similarity and Diagonalizability

Eigenvectors and Eigenvalues of Similar Matrices

**Theorem 8.2.5** Suppose that \( C = P^{-1}AP \) and that \( \lambda \) is an eigenvalue of \( A \) and \( C \).

(a) If \( x \) is an eigenvector of \( C \) corresponding to \( \lambda \), then \( Px \) is an eigenvector of \( A \) corresponding to \( \lambda \).

(b) If \( x \) is an eigenvector of \( A \) corresponding to \( \lambda \), then \( P^{-1}x \) is an eigenvector of \( C \) corresponding to \( \lambda \).

Assume that \( x \) is an eigenvector of \( C \) corresponding to \( \lambda \), so \( x \neq 0 \) and \( Cx = \lambda x \).

If we substitute \( P^{-1}AP \) for \( C \), we obtain

\[
P^{-1}APx = \lambda x
\]

which we can rewrite as

\[
APx = P\lambda x \quad \text{or equivalently,} \quad A(Px) = \lambda(Px)
\]  

(13)

Since \( P \) is invertible and \( x \neq 0 \), it follows that \( Px \neq 0 \).

Thus, the second equation in (13) implies that \( Px \) is an eigenvector of \( A \) corresponding to \( \lambda \).
8.2. Similarity and Diagonalizability

Diagonalization

Diagonal matrices play an important role in many applications because, in many respects, they represent the simplest kinds of linear operators.

Multiplying $\mathbf{x}$ by $D$ has the effect of “scaling” each coordinate of $\mathbf{w}$ (with a sign reversal for negative $d$’s).

$$D\mathbf{x} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1x_1 \\ d_2x_2 \\ \vdots \\ d_nx_n \end{bmatrix}$$

The Diagonalization Problem  Given a square matrix $A$, does there exist an invertible matrix $P$ for which $P^{-1}AP$ is a diagonal matrix, and if so, how does one find such a $P$? If such a matrix $P$ exists, then $A$ is said to be **diagonalizable**, and $P$ is said to **diagonalize** $A$. 

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.2. Similarity and Diagonalizability

Diagonalization

**Theorem 8.2.6** An $n \times n$ matrix $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors.

We will show first that if the matrix $A$ is diagonalizable, then it has $n$ linearly independent eigenvector. The diagonalizability of $A$ implies that there is an invertible matrix $P$ and a diagonal matrix $D$, say

\[
P = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1n} \\
p_{21} & p_{22} & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & p_{nn}
\end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{bmatrix}
\] (14)

such that $P^{-1}AP = D$. If we rewrite this as $AP = PD$ and substitute (14), we obtain

\[
AP = PD = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1n} \\
p_{21} & p_{22} & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & p_{nn}
\end{bmatrix} \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{bmatrix} = \begin{bmatrix}
\lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\
\lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn}
\end{bmatrix}
\] (15)
8.2. Similarity and Diagonalizability

**Diagonalization**

Thus, if we denote the column vectors of $P$ by $p_1, p_2, \ldots, p_n$, then the left side of (15) can be expressed as

$$AP = A[p_1 \ p_2 \ \cdots \ p_n] = [Ap_1 \ Ap_2 \ \cdots \ Ap_n] \quad (16)$$

and the right side of (15) as

$$[\lambda_1 p_1 \ \lambda_2 p_2 \ \cdots \ \lambda_n p_n] \quad (17)$$

It follows from (16) and (17) that

$$Ap_1 = \lambda_1 p_1, \quad Ap_2 = \lambda_2 p_2, \ldots, \quad Ap_n = \lambda_n p_n$$

and it follows from the invertibility of $P$ that $p_1, p_2, \ldots, p_n$ are nonzero, so we have shown that $p_1, p_2, \ldots, p_n$ are eigenvectors of $A$ corresponding to $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively.

Moreover, the invertibility of $P$ also implies that $p_1, p_2, \ldots, p_n$ are linearly independent (Theorem 7.4.4 applied to $P$), so the column vectors of $P$ form a set of $n$ linearly independent eigenvectors of $A$. 
### 8.2. Similarity and Diagonalizability

**Diagonalization**

Conversely, assume that $A$ has $n$ linearly independent eigenvectors, $p_1, p_2, \ldots, p_n$, and that the corresponding eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_n$, so

$$A p_1 = \lambda_1 p_1, \quad A p_2 = \lambda_2 p_2, \ldots, \quad A p_n = \lambda_n p_n$$

If we now form the matrices

$$P = [p_1 \ p_2 \ \cdots \ p_n] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

then we obtain

$$AP = A[p_1 \ p_2 \ \cdots \ p_n] = [Ap_1 \ Ap_2 \ \cdots \ Ap_n] = [\lambda_1 p_1 \ \lambda_2 p_2 \ \cdots \ \lambda_n p_n] = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD$$

However, the matrix $P$ is invertible, since its column vectors are linearly independent, so it follows from this computation that $D = P^{-1}AP$, which shows that $A$ is diagonalizable.
8.2. Similarity and Diagonalizability

A Method for Diagonalizing A Matrix

**Diagonalizing an** $n \times n$ **Matrix with** $n$ **Linearly Independent Eigenvectors**

**Step 1.** Find $n$ linearly independent eigenvectors of $A$, say $p_1, p_2, \ldots, p_n$.

**Step 2.** Form the matrix $P = [p_1 \ p_2 \ \cdots \ p_n]$.

**Step 3.** The matrix $P^{-1}AP$ will be diagonal and will have the eigenvalues corresponding to $p_1, p_2, \ldots, p_n$, respectively, as its successive diagonal entries.
8.2. Similarity and Diagonalizability

A Method for Diagonalizing A Matrix

Example 4

We showed in Example 3 that the matrix

\[
A = \begin{bmatrix}
0 & 0 & -2 \\
1 & 2 & 1 \\
1 & 0 & 3
\end{bmatrix}
\]

has eigenvalues \(\lambda = 1\) and \(\lambda = 2\) and that basis vectors for these eigenspaces are

\[
\mathbf{p}_1 = \begin{bmatrix}
-2 \\
1 \\
1
\end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix}
-1 \\
0 \\
1
\end{bmatrix}, \quad \mathbf{p}_3 = \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\]

\(\lambda = 1\) \quad \(\lambda = 2\)

It is a straightforward matter to show that these three vectors are linearly independent, so \(A\) is diagonalizable and is diagonalized by

\[
P = \begin{bmatrix}
-2 & -1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{bmatrix}
\]
8.2. Similarity and Diagonalizability

A Method for Diagonalizing A Matrix

As a check,

\[
P^{-1}AP = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}
\]

REMARK

There is no preferred order for the columns of a diagonalizing matrix \( P \) – the only effect of changing the order of the column is to change the order in which the eigenvalues appear along the main diagonal of \( D=P^{-1}AP \).

\[
P = [p_3 \quad p_1 \quad p_2] = \begin{bmatrix} 0 & -2 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}
\]
8.2. Similarity and Diagonalizability

A Method for Diagonalizing A Matrix

Example 5

We showed in Example 2 that the matrix

\[ A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ -3 & 5 & 3 \end{bmatrix} \]

has eigenvalues \( \lambda = 2 \) and \( \lambda = 3 \) and that bases for the corresponding eigenspaces are

\[ \mathbf{p}_1 = \begin{bmatrix} \frac{1}{8} \\ -\frac{1}{8} \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \]

\( \lambda = 2 \) \quad \lambda = 3

A is not diagonalizable since all other eigenvectors must be scalar multiples of one of these two.
8.2. Similarity and Diagonalizability

Linearity Independence of Eigenvectors

**Theorem 8.2.7** If \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \) are eigenvectors of a matrix \( A \) that correspond to distinct eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_k \), then the set \( \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \) is linearly independent.

**Theorem 8.2.8** An \( n \times n \) matrix with \( n \) distinct real eigenvalues is diagonalizable.

**Example 6**

The \( 3 \times 3 \) matrix

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
1 & 3 & 0 \\
-3 & 5 & 4 \\
\end{bmatrix}
\]

is diagonalizable, since it has three distinct eigenvalues, \( \lambda=2, \lambda=3, \) and \( \lambda=4 \).

The converse of Theorem 8.2.8 is *false*; that is, it is possible for an \( n \times n \) matrix to be diagonalizable without having \( n \) distinct eigenvalues.
8.2. Similarity and Diagonalizability

Linearly Independence of Eigenvectors

The key to diagonalizability rests with the dimensions of the eigenspaces.

**Theorem 8.2.9** An $n \times n$ matrix $A$ is diagonalizable if and only if the sum of the geometric multiplicities of its eigenvalues is $n$.

Example 7

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
1 & 3 & 0 \\
-3 & 5 & 3
\end{bmatrix}
\quad A = \begin{bmatrix}
0 & 0 & -2 \\
1 & 2 & 1 \\
1 & 0 & 3
\end{bmatrix}
\]

Eigenvalues

- $\lambda = 2$ and $\lambda = 3$
- $\lambda = 1$ and $\lambda = 2$

Geometric multiplicities

- $1$  
- $1$  
- $1$  
- $2$

Diagonalizable

- X
- O
8.2. Similarity and Diagonalizability

Relationship between Algebraic and Geometric Multiplicity

The geometric multiplicity of an eigenvalue cannot exceed its algebraic multiplicity.
For example, if the characteristic polynomial of some $6 \times 6$ matrix $A$ is

$$p(\lambda) = (\lambda - 3)(\lambda - 5)^2(\lambda - 6)^3$$

The eigenspace corresponding to $\lambda=6$ might have dimension 1, 2, or 3, the eigenspace corresponding to $\lambda=5$ might have dimension 1 or 2, and the eigenspace corresponding to $\lambda=3$ must have dimension 1.

**Theorem 8.2.10** If $A$ is a square matrix, then:

(a) The geometric multiplicity of an eigenvalue of $A$ is less than or equal to its algebraic multiplicity.

(b) $A$ is diagonalizable if and only if the geometric multiplicity of each eigenvalue of $A$ is the same as its algebraic multiplicity.
8.2. Similarity and Diagonalizability

A Unifying Theorem on Diagonalizability

**Theorem 8.2.11**  If $A$ is an $n \times n$ matrix, then the following statements are equivalent.

(a) $A$ is diagonalizable.
(b) $A$ has $n$ linearly independent eigenvectors.
(c) $\mathbb{R}^n$ has a basis consisting of eigenvectors of $A$.
(d) The sum of the geometric multiplicities of the eigenvalues of $A$ is $n$.
(e) The geometric multiplicity of each eigenvalue of $A$ is the same as the algebraic multiplicity.

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Orthogonal Similarity

Two $n \times n$ matrices $A$ and $C$ are said to be similar if there exists an invertible matrix $P$ such that $C = P^{-1}AP$.

**Definition 8.3.1** If $A$ and $C$ are square matrices with the same size, then we say that $C$ is orthogonally similar to $A$ if there exists an orthogonal matrix $P$ such that $C = P^TAP$.

If $C$ is orthogonally similar to $A$, then $A$ is orthogonally similar to $C$, so we will usually say that $A$ and $C$ are orthogonally similar.

**Theorem 8.3.2** Two matrices are orthogonally similar if and only if there exist orthonormal bases with respect to which the matrices represent the same linear operator.
Orthogonal Similarity

The Orthogonal Diagonalization Problem  Given a square matrix $A$, does there exist an orthogonal matrix $P$ for which $P^TAP$ is a diagonal matrix, and if so, how does one find such a $P$? If such a matrix $P$ exists, then $A$ is said to be orthogonally diagonalizable, and $P$ is said to orthogonally diagonalize $A$. 
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Orthogonal Similarity

Suppose that

\[ D = P^TAP \]  \hspace{1cm} (1)  

where \( P \) is orthogonal and \( D \) is diagonal. Since \( P^TP = PP^T = I \), we can rewrite (1) as

\[ A = PDP^T \]

Transposing both sides of this equation and using the fact that \( D^T = D \) yields

\[ A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A \]

which shows that an orthogonally diagonalizable matrix must be symmetric.
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Orthogonal Similarity

**Theorem 8.3.3** An $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if there exists an orthonormal set of $n$ eigenvectors of $A$.

The orthogonal diagonalizability of $A$ implies that there exists an orthogonal matrix $P$ and a diagonal matrix $D$ such that $P^TAP = D$.

However, since the column vectors of an orthogonal matrix are orthonormal, and since the column vectors of $P$ are eigenvectors of $A$ (see the proof of Theorem 8.2.6), we have established that the column vectors of $P$ form an orthonormal set of $n$ eigenvectors of $A$.

Conversely, assume that there exists an orthonormal set $\{p_1, p_2, \ldots, p_n\}$ of $n$ eigenvectors of $A$. We showed in the proof of Theorem 8.2.6 that the matrix

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

diagonalizes $A$. However, in this case $P$ is an orthogonal matrix, since its column vectors are orthonormal.

Thus, $P$ orthogonally diagonalizes $A$. 

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
Orthogonal Similarity

**Theorem 8.3.4**

(a) A matrix is orthogonally diagonalizable if and only if it is symmetric.

(b) If A is a symmetric matrix, then eigenvectors from different eigenspaces are orthogonal.

(b) Let \( v_1 \) and \( v_2 \) be eigenvectors corresponding to distinct eigenvalues \( \lambda_1 \) and \( \lambda_2 \), respectively. The proof that \( v_1 \cdot v_2 = 0 \) will be facilitated by using Formula (26) of Section 3.1 to write

\[
\lambda_1 (v_1 \cdot v_2) = (\lambda_1 v_1) \cdot v_2
\]

as the matrix product \( (\lambda_1 v_1)^T v_2 \). The rest of the proof now consists of manipulating this expression in the right way:

\[
\lambda_1 (v_1 \cdot v_2) = (\lambda_1 v_1)^T v_2 = (Av_1)^T v_2 = (v_1^T A^T) v_2
\]

[\( v_1 \) is an eigenvector corresponding to \( \lambda_1 \)]

\[
= (v_1^T A) v_2
\]

[symmetry of \( A \)]

\[
= v_1^T (Av_2)
\]

\[
= v_1^T (\lambda_2 v_2)
\]

[\( v_2 \) is an eigenvector corresponding to \( \lambda_2 \)]

\[
= \lambda_2 v_1^T v_2
\]

\[
= \lambda_2 (v_1 \cdot v_2)
\]

[Formula (26) of Section 3.1]

This implies that \( (\lambda_1 - \lambda_2)(v_1 \cdot v_2) = 0 \), so \( v_1 \cdot v_2 = 0 \) as a result of the fact that \( \lambda_1 \neq \lambda_2 \).
8.3. Orthogonal Diagonalizability; Functions of a Matrix

A Method for Orthogonally Diagonalizing A Symmetric Matrix

Orthogonally Diagonalizing an \( n \times n \) Symmetric Matrix

**Step 1.** Find a basis for each eigenspace of \( A \).

**Step 2.** Apply the Gram–Schmidt process to each of these bases to produce orthonormal bases for the eigenspaces.

**Step 3.** Form the matrix \( P = [p_1 \ p_2 \ \cdots \ p_n] \) whose columns are the vectors constructed in Step 2. The matrix \( P \) will orthogonally diagonalize \( A \), and the eigenvalues on the diagonal of \( D = P^TAP \) will be in the same order as their corresponding eigenvectors in \( P \).
8.3. Orthogonal Diagonalizability; Functions of a Matrix

A Method for Orthogonally Diagonalizing A Symmetric Matrix

Example 1

Find a matrix $P$ that orthogonally diagonalizes the symmetric matrix

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$$

The characteristic equation of $A$ is

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 4 & -2 & -2 \\ -2 & \lambda - 4 & -2 \\ -2 & -2 & \lambda - 4 \end{bmatrix} = (\lambda - 2)^2(\lambda - 8) = 0$$

Thus, the eigenvalues of $A$ are $\lambda=2$ and $\lambda=8$. Using the method given in Example 3 of Section 8.2, it can be shown that the vectors

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

form a basis for the eigenspace corresponding to $\lambda=2$ and that
A Method for Orthogonally Diagonalizing A Symmetric Matrix

**Example 1**

\[ v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \]

is a basis for the eigenspace corresponding to \( \lambda = 8 \). Applying the Gram-Schmidt process to the bases \( \{v_1, v_2\} \) and \( \{v_3\} \) yields the orthonormal bases \( \{u_1, u_2\} \) and \( \{u_3\} \), where

\[ u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix} \quad \text{and} \quad u_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} \]

Thus, \( A \) is orthogonally diagonalized by the matrix

\[
P = \begin{bmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}
\end{bmatrix}
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

A Method for Orthogonally Diagonalizing A Symmetric Matrix

Example 1

As a check,

\[ P^TAP = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix} \]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Spectral Decomposition

If \( A \) is a symmetric matrix that is orthogonally diagonalized by

\[
P = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}
\]

and if \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are the eigenvalues of \( A \) corresponding to \( u_1, u_2, \ldots, u_n \), then we know that \( D = P^T A P \), where \( D \) is a diagonal matrix with the eigenvalues in the diagonal positions.

It follows from this that the matrix \( A \) can be expressed as

\[
A = PDP^T = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} = [\lambda_1 u_1 \quad \lambda_2 u_2 \quad \cdots \quad \lambda_n u_n]
\]

Multiplying out using the column-row rule (Theorem 3.8.1), we obtain the formula

\[
A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T
\]

which is called a spectral decomposition of \( A \) or an eigenvalue decomposition of \( A \) (sometimes abbreviated as the EVD of \( A \)).
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Spectral Decomposition

Example 2

The matrix

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

has eigenvalues \( \lambda_1 = -3 \) and \( \lambda_2 = 2 \) with corresponding eigenvectors

\[ \mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \]

Normalizing these basis vectors yields

\[ \mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \]

so a spectral decomposition of \( A \) is

\[ \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T = (-3) \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} + (2) \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Spectral Decomposition

Example 2

\[
\begin{pmatrix}
\frac{1}{5} & -2 \\
-\frac{2}{5} & \frac{4}{5}
\end{pmatrix}
+ (2)
\begin{pmatrix}
\frac{4}{5} & \frac{2}{5} \\
\frac{2}{5} & \frac{1}{5}
\end{pmatrix}
\] (6)

where the 2×2 matrices on the right are the standard matrices for the orthogonal projections onto the eigenspaces.

Now let us see what this decomposition tells us about the image of the vector \(x=(1,1)\) under multiplication by \(A\). Writing \(x\) in column form, it follows that

\[
Ax = \begin{pmatrix}
1 & 2 \\
2 & -2
\end{pmatrix}
\begin{pmatrix}
1 \\
1
\end{pmatrix}
= \begin{pmatrix}
3 \\
0
\end{pmatrix}
\] (7)

and from (6) that

\[
Ax = \begin{pmatrix}
1 & 2 \\
2 & -2
\end{pmatrix}
\begin{pmatrix}
1 \\
1
\end{pmatrix}
= (-3)
\begin{pmatrix}
\frac{1}{5} & -\frac{2}{5} \\
-\frac{2}{5} & \frac{4}{5}
\end{pmatrix}
\begin{pmatrix}
1 \\
1
\end{pmatrix}
+ (2)
\begin{pmatrix}
\frac{4}{5} & \frac{2}{5} \\
\frac{2}{5} & \frac{1}{5}
\end{pmatrix}
\begin{pmatrix}
1 \\
1
\end{pmatrix}
\] (8)

\[
= (-3)
\begin{pmatrix}
-\frac{1}{5} \\
\frac{2}{5}
\end{pmatrix}
+ (2)
\begin{pmatrix}
\frac{6}{5} \\
\frac{3}{5}
\end{pmatrix}
= \begin{pmatrix}
\frac{3}{5} \\
-\frac{6}{5}
\end{pmatrix}
+ \begin{pmatrix}
\frac{12}{5} \\
\frac{6}{5}
\end{pmatrix}
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Spectral Decomposition

Example 2

It follows from (7) that the image of (1,1) under multiplication by $A$ is (3,0), and it follows from (8) that this image can also be obtained by projecting (1,1) onto the eigenspaces corresponding to $\lambda_1=-3$ and $\lambda_2=2$ to obtain vectors $(-\frac{1}{5}, \frac{2}{5})$ and $(\frac{6}{5}, \frac{3}{5})$, then scaling by the eigenvalues to obtain $(\frac{3}{5}, -\frac{6}{5})$ and $(\frac{12}{5}, \frac{6}{5})$, and then adding these vectors.

\[
\begin{align*}
\text{x} &= (1,1) \\
A\text{x} &= (3,0) \\
\lambda &= 2 \\
\left(-\frac{1}{5}, \frac{2}{5}\right) &\quad \left(\frac{6}{5}, \frac{3}{5}\right) \\
\lambda &= 3 \\
\left(\frac{3}{5}, -\frac{6}{5}\right) &\quad \left(\frac{12}{5}, \frac{6}{5}\right)
\end{align*}
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Powers of A Diagonalizable Matrix

Suppose that $A$ is an $n \times n$ matrix and $P$ is an invertible $n \times n$ matrix. Then

$$(P^{-1}AP)^2 = (P^{-1}AP)(P^{-1}AP) = P^{-1}APP^{-1}AP = P^{-1}AIA = P^{-1}A^2P$$

and more generally, if $k$ is any positive integer, then

$$(P^{-1}AP)^k = P^{-1}A^kP \quad (9)$$

In particular, if $A$ is diagonalizable and $P^{-1}AP = D$ is a diagonal matrix, then it follows from (9) that

$$P^{-1}A^kP = D^k$$

which we can rewrite as

$$A^k = PD^kP^{-1} \quad (11)$$
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Powers of A Diagonalizable Matrix

Example 3

Use Formula (11) to find $A^{13}$ for the diagonalizable matrix

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

We showed in Example 4 of Section 8.2 that

$$P^{-1}AP = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus,

$$A^{13} = \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1^{13} & 0 & 0 \\ 0 & 2^{13} & 0 \\ 0 & 0 & 2^{13} \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -8190 & 0 & -16382 \\ 8191 & 8192 & 8191 \\ 8191 & 0 & 16383 \end{bmatrix}$$
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Powers of A Diagonalizable Matrix

In the special case where $A$ is a symmetric matrix with a spectral decomposition

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$$

the matrix

$$P = [u_1 \quad u_2 \quad \cdots \quad u_n]$$

orthogonally diagonalizes $A$, so (11) can be expressed as

$$A^k = PD^k P^T$$

This equation can be written as

$$A^k = \lambda_1^k u_1 u_1^T + \lambda_2^k u_2 u_2^T + \cdots + \lambda_n^k u_n u_n^T$$

from which it follows that $A^k$ is a symmetric matrix whose eigenvalues are the $k$th powers of the eigenvalues of $A$. 
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Cayley-Hamilton Theorem

**Theorem 8.3.5 (Cayley–Hamilton Theorem)** Every square matrix satisfies its characteristic equation; that is, if $A$ is an $n \times n$ matrix whose characteristic equation is

$$\lambda^n + c_1 \lambda^{n-1} + \cdots + c_n = 0$$

then

$$A^n + c_1 A^{n-1} + \cdots + c_n I = 0 \quad (14)$$

The Cayley-Hamilton theorem makes it possible to express all positive integer powers of an $n \times n$ matrix $A$ in terms of $I, A, \ldots, A^{n-1}$ by solving (14) of $A^n$.

In the case where $A$ is invertible, it also makes it possible to express $A^{-1}$ (and hence all negative powers of $A$) in terms of $I, A, \ldots, A^{n-1}$ by rewriting (14) as

$$A \left( -\frac{1}{c_n} A^{n-1} - \frac{c_1}{c_n} A^{n-2} - \cdots - \frac{c_{n-1}}{c_n} I \right) = I$$

, from which it follows that $A^{-1}$ is the parenthetical expression on the left.
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Cayley-Hamilton Theorem

**Example 4**

We showed in Example 2 of Section 8.2 that the characteristic polynomial of

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
1 & 3 & 0 \\
-3 & 5 & 3 \\
\end{bmatrix}
\]

is

\[
p(\lambda) = (\lambda - 2)(\lambda - 3)^2 = \lambda^3 - 8\lambda^2 + 21\lambda - 18
\]

so the Cayley-Hamilton theorem implies that

\[
A^3 - 8A^2 + 21A - 18I = 0
\]  \hspace{1cm} (16)

This equation can be used to express \( A^3 \) and all higher powers of \( A \) in terms of \( I \), \( A \), and \( A^2 \). For example,

\[
A^3 = 8A^2 - 21A + 18I
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Cayley-Hamilton Theorem

Example 4

and using this equation we can write

\[ A^4 = AA^3 = 8A^3 - 21A^2 + 18A = 8(8A^2 - 21A + 18I) - 21A^2 + 18A \]
\[ = 43A^2 - 150A + 144I \]

Equation (16) can also be used to express \( A^{-1} \) as a polynomial in \( A \) by rewriting it as

\[ A(A^2 - 8A + 21I) = 18I \]

from which it follows that

\[ A^{-1} = \frac{1}{18}(A^2 - 8A + 21I) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & \frac{1}{3} & 0 \\ \frac{7}{9} & -\frac{5}{9} & \frac{1}{3} \end{bmatrix} \]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

In Section 3.2 we defined polynomial functions of square matrices. Recall from that discussion that if $A$ is an $n \times n$ matrix and

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

then the matrix $p(A)$ is defined as

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m$$
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

Other functions of square matrices can be defined using power series.

For example, if the function $f$ is represented by its Maclaurin series

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^m(0)}{m!}x^m + \cdots$$

on some interval, then we define $f(A)$ to be

$$f(A) = f(0)I + f'(0)A + \frac{f''(0)}{2!}A^2 + \cdots + \frac{f^m(0)}{m!}A^m + \cdots \quad (18)$$

where we interpret this to mean that the $ij$th entry of $f(A)$ is the sum of the series of the $ij$th entries of the terms on the right.
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

In the special case where \( A \) is a diagonal matrix, say

\[
A = \begin{bmatrix}
    d_1 & 0 & \cdots & 0 \\
    0 & d_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & d_n
\end{bmatrix}
\]

and \( f \) is defined at the point \( d_1, d_2, \ldots, d_k \), each matrix on the right side of (18) is diagonal, and hence so is \( f(A) \).

In this case, equating corresponding diagonal entries on the two sides of (18) yields

\[
(f(A))_{kk} = f(0) + f'(0)d_k + \frac{f''(0)}{2!}d_k^2 + \cdots + \frac{f^m(0)}{m!}d_k^m + \cdots = f(d_k)
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of a Matrix

Thus, we can avoid the series altogether in the diagonal case and compute $f(A)$ directly as

$$f(A) = \begin{bmatrix} f(d_1) & 0 & \cdots & 0 \\ 0 & f(d_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(d_n) \end{bmatrix}$$

For example, if

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

then

$$e^A = \begin{bmatrix} e & 0 & 0 \\ 0 & e^3 & 0 \\ 0 & 0 & e^{-2} \end{bmatrix}$$
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

If $A$ is an $n \times n$ diagonalizable matrix and $P^{-1}AP = D$, where

$$D = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{bmatrix}$$

then (10) and (18) suggest that

$$P^{-1}f(A)P = f(0)I + f'(0)(P^{-1}AP) + \frac{f''(0)}{2!}(P^{-1}A^2P) + \cdots + \frac{f^{m}(0)}{m!}(P^{-1}A^mP) + \cdots$$

$$= f(0)I + f'(0)D + \frac{f''(0)}{2!}D^2 + \cdots + \frac{f^{m}(0)}{m!}D^m + \cdots$$

$$= f(D)$$

This tells us that $f(A)$ can be expressed as

$$f(A) = Pf(D)P^{-1}$$
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

**Theorem 8.3.6** Suppose that $A$ is an $n \times n$ diagonalizable matrix that is diagonalized by $P$ and that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $A$ corresponding to the successive column vectors of $P$. If $f$ is a real-valued function whose Maclaurin series converges on some interval containing the eigenvalues of $A$, then

$$f(A) = P \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} P^{-1}$$

(21)
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

Example 5

Find $e^{tA}$ for the diagonalizable matrix

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

We showed in Example 4 of Section 8.2 that

$$P^{-1}AP = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$
Exponential of A Matrix

Example 5

so applying Formula (21) with \( f(A) = e^{tA} \) implies that

\[
e^{tA} = P \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{2t} \end{bmatrix} P^{-1} = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}
\]

\[
= \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ e^{2t} - e^t & e^{2t} & e^{2t} - e^t \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix}
\]
8.3. Orthogonal Diagonalizability; Functions of a Matrix

Exponential of A Matrix

In the special case where $A$ is a symmetric matrix with a spectral decomposition
\[ A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T \]
the matrix
\[ P = [u_1 \quad u_2 \quad \cdots \quad u_n] \]
orthogonally diagonalizes $A$, so (20) can be expressed as
\[ f(A) = Pf(D)P^T \]
This equation can be written as
\[ f(A) = f(\lambda_1)u_1 u_1^T + f(\lambda_2)u_2 u_2^T + \cdots + f(\lambda_n)u_n u_n^T \]
, which tells us that $f(A)$ is a symmetric matrix whose eigenvalues can be obtained by evaluating $f$ at the eigenvalues of $A$. 
Diagonalization and Linear Systems

Assume that we are solving a linear system $Ax = b$.

Suppose that $A$ is diagonalizable and $P^{-1}AP = D$. If we define a new vector $y = P^{-1}x$, and if we substitute

$$x = Py \quad (23)$$

in $Ax = b$, then we obtain a new linear system $APy = b$ in the unknown $y$. Multiplying both sides of this equation by $P^{-1}$ and using the fact that $P^{-1}AP = D$ yields

$$Dy = P^{-1}b$$

Since this system has a diagonal coefficient matrix, the solution for $y$ can be read off immediately, and the vector $x$ can then be computed using (23).

Many algorithms for solving large-scale linear systems are based on this idea. Such algorithms are particularly effective in cases in which the coefficient matrix can be orthogonally diagonalized since multiplication by orthogonal matrices does not magnify roundoff error.
The Nondiagonalizable Case

In cases where $A$ is not diagonalizable it is still possible to achieve considerable simplification in the form of $P^{-1}AP$ by choosing the matrix $P$ appropriately.

**Theorem 8.3.7 (Schur’s Theorem)** If $A$ is an $n \times n$ matrix with real entries and real eigenvalues, then there is an orthogonal matrix $P$ such that $P^TAP$ is an upper triangular matrix of the form

$$ P^TAP = \begin{bmatrix}
\lambda_1 & \times & \times & \cdots & \times \\
0 & \lambda_2 & \times & \cdots & \times \\
0 & 0 & \lambda_3 & \cdots & \times \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_n
\end{bmatrix} $$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the matrix $A$ repeated according to multiplicity.
8.3. Orthogonal Diagonalizability; Functions of a Matrix

The Nondiagonalizable Case

**Theorem 8.3.8 (Hessenberg’s Theorem)** Every square matrix with real entries is orthogonally similar to a matrix in upper Hessenberg form; that is, if $A$ is an $n \times n$ matrix, then there is an orthogonal matrix $P$ such that $P^TAP$ is a matrix of the form

$$
P^TAP = \begin{bmatrix}
\times & \times & \cdots & \times & \times & \times \\
\times & \times & \cdots & \times & \times & \times \\
0 & \times & \cdots & \times & \times & \times \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \times & \times & \times \\
0 & 0 & \cdots & 0 & \times & \times \\
\end{bmatrix}
$$

(26)

In many numerical $LU$- and $QR$-algorithms in initial matrix is first converted to upper Hessenberg form, thereby reducing the amount of computation in the algorithm itself.
8.4. Quadratic Forms

Definition of Quadratic Form

A **linear form** on $R^n$

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

A **quadratic form** on $R^n$

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2 + \text{(all possible terms } a_kx_ix_j \text{ in which } x_i \text{ and } x_j \text{ are distinct)}$$

The term of the form $a_kx_ix_j$ are called **cross product terms**.

A general quadratic form on $R^2$

$$a_1x_1^2 + a_2x_2^2 + 2a_3x_1x_2 \quad \Rightarrow \quad a_1x_1^2 + a_2x_2^2 + 2a_3x_1x_2 = [x_1 \quad x_2] \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix} [x_1 \quad x_2] = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

A general quadratic form on $R^3$

$$a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + 2a_4x_1x_2 + 2a_5x_1x_3 + 2a_6x_2x_3$$

$$a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + 2a_4x_1x_2 + 2a_5x_1x_3 + 2a_6x_2x_3 = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} a_1 & a_4 & a_5 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{bmatrix} [x_1 \quad x_2 \quad x_3] = \mathbf{x}^T \mathbf{A} \mathbf{x}$$
8.4. Quadratic Forms

Definition of Quadratic Form

In general, if $A$ is a symmetric $n \times n$ matrix and $x$ is an $n \times 1$ column vector of variables, then we call the function

$$Q_A(x) = x^T Ax$$

(3)

the \textit{quadratic form associated with $A$}.

When convenient, (3) can be expressed in dot product notation as

$$Q_A(x) = x^T Ax = x^T (Ax) = x \cdot Ax = A x \cdot x$$
8.4. Quadratic Forms

Definition of Quadratic Form

In the case where \( A \) is a diagonal matrix, the quadratic form \( Q_A \) has no cross product terms; for example, if \( A \) is the \( n \times n \) identity matrix, then

\[
Q_A(x) = x^T I x = x^T x = \|x\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2
\]

and if \( A \) has diagonal entries \( \lambda_1, \lambda_2, \ldots, \lambda_n \), then

\[
Q_A(x) = x^T A x = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{bmatrix} [x_1 \quad x_2 \quad \cdots \quad x_n]^T = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2
\]
8.4. Quadratic Forms

Definition of Quadratic Form

Example 1

In each part, express the quadratic form in the matrix notation $\mathbf{x}^T A \mathbf{x}$, where $A$ is symmetric.

(a) $2x^2 + 6xy - 5y^2$  
(b) $x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_2$

$$2x^2 + 6xy - 5y^2 = [\mathbf{x} \quad \mathbf{y}] \begin{bmatrix} 2 & 3 \\ 3 & -5 \end{bmatrix} [\mathbf{x} \quad \mathbf{y}]$$

$$x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_3 = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3] \begin{bmatrix} 1 & 2 & -1 \\ 2 & 7 & 4 \\ -1 & 4 & -3 \end{bmatrix} [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3]$$
8.4. Quadratic Forms

Change of Variable in A Quadratic Form

We can simplify the quadratic form $x^TAx$ by making a substitution

$$x = Py$$

that expresses the variable $x_1, x_2, \ldots, x_n$ in terms of new variable $y_1, y_2, \ldots, y_n$.

If $P$ is invertible, then we call (5) a *change of variable*, and if $P$ is orthogonal, we call (5) an *orthogonal change of variable*.

If we make the change of variable $x=Py$ in the quadratic form $x^TAx$, then we obtain

$$x^TAx = (Py)^TA(Py) = y^TP^TAy = y^T(P^TAp)y$$

The matrix $B=P^TAp$ is symmetric, so the effect of the change of variable is to produce a new quadratic form $y^TBy$ in the variable $y_1, y_2, \ldots, y_n$. 
8.4. Quadratic Forms

Change of Variable in A Quadratic Form

In particular, if we choose $P$ to orthogonally diagonalize $A$, then the new quadratic form will be $y^TDy$, where $D$ is a diagonal matrix with the eigenvalues of $A$ on the main diagonal; that is

$$x^TAx = y^TDy = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

**Theorem 8.4.1 (The Principal Axes Theorem)** If $A$ is a symmetric $n \times n$ matrix, then there is an orthogonal change of variable that transforms the quadratic form $x^TAx$ into a quadratic form $y^TDy$ with no cross product terms. Specifically, if $P$ orthogonally diagonalizes $A$, then making the change of variable $x = Py$ in the quadratic form $x^TAx$ yields the quadratic form

$$x^TAx = y^TDy = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $A$ corresponding to the eigenvectors that form the successive columns of $P$. 

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.4. Quadratic Forms

Change of Variable in A Quadratic Form

Example 2

Find an orthogonal change of variable that eliminates the cross product terms in the quadratic form \( Q = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3 \), and express \( Q \) in terms of the new variables.

\[
Q = x^T A x = [x_1 \quad x_2 \quad x_3] \begin{bmatrix}
1 & -2 & 0 \\
-2 & 0 & 2 \\
0 & 2 & -1
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
\]

The characteristic equation of the matrix \( A \) is

\[
\begin{vmatrix}
\lambda - 1 & 2 & 0 \\
2 & \lambda & -2 \\
0 & -2 & \lambda + 1
\end{vmatrix} = \lambda^3 - 9\lambda = \lambda(\lambda + 3)(\lambda - 3) = 0
\]

so the eigenvalues are \( \lambda = 0, -3, 3. \) the orthonormal bases for the three eigenspaces are

\[
\lambda = 0: \begin{bmatrix}
\frac{2}{3} \\
\frac{1}{3} \\
\frac{2}{3}
\end{bmatrix}, \quad \lambda = -3: \begin{bmatrix}
-\frac{1}{3} \\
-\frac{2}{3} \\
\frac{2}{3}
\end{bmatrix}, \quad \lambda = 3: \begin{bmatrix}
\frac{2}{3} \\
\frac{1}{3} \\
\frac{2}{3}
\end{bmatrix}
\]
8.4. Quadratic Forms

Change of Variable in A Quadratic Form

Example 2

Thus, a substitution \( x = Py \) that eliminates the cross product terms is

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix} =
\begin{bmatrix}
  \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\
  \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\
  \frac{2}{3} & \frac{2}{3} & \frac{1}{3}
\end{bmatrix}
\begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix}
\]

This produces the new quadratic form

\[
Q = y^T(P^TAP)y = \begin{bmatrix}
  y_1 & y_2 & y_3
\end{bmatrix}
\begin{bmatrix}
  0 & 0 & 0 \\
  0 & -3 & 0 \\
  0 & 0 & 3
\end{bmatrix}
\begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix}
= -3y_2^2 + 3y_3^2
\]
8.4. Quadratic Forms

Quadratic Forms in Geometry

Recall that a conic section or conic is a curve that results by cutting a double-napped cone with a plane. The most important conic sections are ellipse, hyperbolas, and parabolas, which occur when the cutting plane does not pass through the vertex.

If the cutting plane passes through the vertex, then the resulting intersection is called a degenerate conic. The possibilities are a point, a pair of intersecting lines, or a single line.
8.4. Quadratic Forms

Quadratic Forms in Geometry

\[ ax^2 + 2bxy + cy^2 + dx + ey + f = 0 \] conic section
\[ ax^2 + 2bxy + cy^2 + f = 0 \] central conic
\[ ax^2 + cy^2 + f = 0 \] central conic in **standard position**
\[ a'x^2 + b'y^2 = 1 \] **standard forms** of nondegenerate central conic

\[
\begin{align*}
\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} &= 1 \\
&\text{ } (\alpha \geq \beta > 0) \\
\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} &= 1 \\
&\text{ } (\beta \geq \alpha > 0) \\
\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} &= 1 \\
&\text{ } (\alpha > 0, \beta > 0) \\
\frac{y^2}{\beta^2} - \frac{x^2}{\alpha^2} &= 1 \\
&\text{ } (\alpha > 0, \beta > 0)
\end{align*}
\]
8.4. Quadratic Forms

Quadratic Forms in Geometry

A central conic whose equation has a cross product term results by rotating a conic in standard position about the origin and hence is said to \textit{rotated out of standard position}.

![Graph of a central conic](graph.png)

Quadratic forms on $\mathbb{R}^3$ arise in the study of geometric objects called \textit{quadratic surfaces} (or \textit{quadrics}). The most important surfaces of this type have equations of the form

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + g = 0$$

in which $a$, $b$, and $c$ are not all zero. These are called \textit{central quadrics}.
8.4. Quadratic Forms

Identifying Conic Sections

It is an easy matter to identify central conics in standard position by matching the equation with one of the standard forms.

For example, the equation

$$9x^2 + 16y^2 - 144 = 0$$

can be rewritten as

$$\frac{x^2}{16} + \frac{y^2}{9} = 1$$

which, by comparison with Table 8.4.1, is the ellipse shown in Figure 8.4.3.
8.4. Quadratic Forms

Identifying Conic Sections

If a central conic is rotated out of standard position, then it can be identified by first rotating
the coordinate axes to put it in standard position and then matching the resulting equation with
one of the standard forms in Table 8.4.1.

\[ ax^2 + 2bxy + cy^2 = k \]

\[ \mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = k \quad (12) \]

To rotate the coordinate axes, we need to make an orthogonal change of variable

\[ \mathbf{x} = P \mathbf{x}' \]

in which \( \det(P)=1 \), and if we want this rotation to eliminate the cross product term, we must
choose \( P \) to orthogonally diagonalize \( A \). If we make a change of variable with these two
properties, then in the rotated coordinate system Equation (12) will become

\[ \mathbf{x}'^T D \mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = k \quad (13) \]

where \( \lambda_1 \) and \( \lambda_2 \) are the eigenvalues of \( A \). the conic can now be identified by writing (13) in
the form

\[ \lambda_1 x'^2 + \lambda_2 y'^2 = k \]
Identifying Conic Sections

The first column vector of $P$, which is a unit eigenvector corresponding to $\lambda_1$, is along the positive $x'$-axis; and the second column vector of $P$, which is eigenvector corresponding to $\lambda_2$, is a unit vector along the $y'$-axis. These are called the **principal axes** of the ellipse.

Also, since $P$ is the transition matrix from $x'y'$-coordinates to $xy$-coordinates, it follows from Formula (29) of Section 7.11 that the matrix $P$ can be expressed in terms of the rotating angle $\theta$ as

$$P = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$
8.4. Quadratic Forms

Identifying Conic Sections

Example 3

(a) Identify the conic whose equation is \(5x^2 - 4xy + 8y^2 - 36 = 0\) by rotating the \(xy\)-axes to put the conic in standard position.

(b) Find the angle \(\theta\) through which you rotated the \(xy\)-axes in part (a).

\[
\mathbf{x}^T \mathbf{A} \mathbf{x} = 36 \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}
\]

The characteristic polynomial of \(A\) is

\[
\begin{vmatrix} \lambda - 5 & 2 \\ 2 & \lambda - 8 \end{vmatrix} = (\lambda - 4)(\lambda - 9)
\]

\[
\lambda = 4: \begin{bmatrix} 2 \\ \sqrt{5} \end{bmatrix}, \quad \lambda = 9: \begin{bmatrix} -1 \\ \sqrt{5} \end{bmatrix}
\]

Thus, \(A\) is orthogonally diagonalized by

\[
P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}
\]

\(\det(P) = 1\). If \(\det(P) = -1\), then we would have exchanged the columns to reverse the sign.
8.4. Quadratic Forms

Identifying Conic Sections

Example 3

The equation of the conic in the $x'y'$-coordinate system is

$$ [x' \quad y'] \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} [x' \quad y'] = 36 $$

which we can write as

$$ 4x'^2 + 9y'^2 = 36 \quad \text{or} \quad \frac{x'^2}{9} + \frac{y'^2}{4} = 1 $$

$$ P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} $$

$$ \cos \theta = \frac{2}{\sqrt{5}}, \quad \sin \theta = \frac{1}{\sqrt{5}}, \quad \tan \theta = \frac{\sin \theta}{\cos \theta} = \frac{1}{2} $$

Thus, $\theta = \tan^{-1} \frac{1}{2} \approx 26.6^\circ$
8.4. Quadratic Forms

Positive Definite Quadratic Forms

**Definition 8.4.2** A quadratic form $x^TAx$ is said to be

*positive definite* if $x^TAx > 0$ for $x \neq 0$

*negative definite* if $x^TAx < 0$ for $x \neq 0$

*indefinite* if $x^TAx$ has both positive and negative values
8.4. Quadratic Forms

Positive Definite Quadratic Forms

**Theorem 8.4.3** If $A$ is a symmetric matrix, then:

(a) $x^T A x$ is positive definite if and only if all eigenvalues of $A$ are positive.

(b) $x^T A x$ is negative definite if and only if all eigenvalues of $A$ are negative.

(c) $x^T A x$ is indefinite if and only if $A$ has at least one positive eigenvalue and at least one negative eigenvalue.

It follows from the principal axes theorem (Theorem 8.4.1) that there is an orthogonal change of variable $x = Py$ for which

$$x^T A x = y^T D y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

Moreover, it follows from the invertibility of $P$ that $y \neq 0$ if and only if $x \neq 0$, so the values of $x^T A x$ for $x \neq 0$ are the same as the values of $y^T D y$ for $y \neq 0$. 
8.4. Quadratic Forms

Classifying Conic Sections Using Eigenvalues

\[ x^T B x = k \]

\[ k \neq 1 \]

\[ x^T A x = 1 \]

\[ \lambda_1 x^2 + \lambda_2 y^2 = 1 \]

- an ellipse if \( \lambda_1 > 0 \) and \( \lambda_2 > 0 \)
- no graph if \( \lambda_1 < 0 \) and \( \lambda_2 < 0 \)
- a hyperbola if \( \lambda_1 \) and \( \lambda_2 \) have opposite signs

**Theorem 8.4.4** If \( A \) is a symmetric \( 2 \times 2 \) matrix, then:

(a) \( x^T A x = 1 \) represents an ellipse if \( A \) is positive definite.
(b) \( x^T A x = 1 \) has no graph if \( A \) is negative definite.
(c) \( x^T A x = 1 \) represents a hyperbola if \( A \) is indefinite.
8.4. Quadratic Forms

Identifying Positive Definite Matrices

**Theorem 8.4.5** A symmetric matrix $A$ is positive definite if and only if the determinant of every principal submatrix is positive.

It can be used to determine whether a symmetric matrix is positive definite without finding the eigenvalues.

**Example 5**

The matrix

$$A = \begin{bmatrix} 2 & -1 & -3 \\ -1 & 2 & 4 \\ -3 & 4 & 9 \end{bmatrix}$$

is positive definite since the determinants

$$|2| = 2, \quad \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3, \quad \begin{vmatrix} 2 & -1 & -3 \\ -1 & 2 & 4 \\ -3 & 4 & 9 \end{vmatrix} = 1$$

Thus, we are guaranteed that all eigenvalues of $A$ are positive and $x^T A x > 0$ for $x \neq 0$. 
8.4. Quadratic Forms

Identifying Positive Definite Matrices

**Theorem 8.4.6** If $A$ is a symmetric matrix, then the following statements are equivalent.

(a) $A$ is positive definite.

(b) There is a symmetric positive definite matrix $B$ such that $A = B^2$.

(c) There is an invertible matrix $C$ such that $A = C^T C$.

(a) $\Rightarrow$ (b)

Since $A$ is symmetric, it is orthogonally diagonalizable. This means that there is an orthogonal matrix $P$ such that $P^T A P = D$, where $D$ is a diagonal matrix whose entries are the eigenvalues of $A$.

Moreover, since $A$ is positive definite, its eigenvalues are positive, so we can write $D$ as $D = D_1^2$, where $D_1$ is the diagonal matrix whose entries are the square roots of the eigenvalues of $A$. Thus, we have $P^T A P = D_1^2$, which we can rewrite as

$$A = P D_1^2 P^T = P D_1 D_1 P^T = P D_1 P^T P D_1 P^T = (P D_1 P^T)(P D_1 P^T) = B^2$$

where $B = P D_1 P^T$. 
8.4. Quadratic Forms

**Identifying Positive Definite Matrices**

The eigenvalues of $B$ are the same as the eigenvalues of $D_1$, since eigenvalues are a similarity invariant and $B$ is similar to $D_1$.

Thus, the eigenvalues of $B$ are positive, since they are the square roots of the eigenvalues of $A$.

(b) $\Rightarrow$ (c)

Assume that $A=B^2$, where $B$ is symmetric and positive definite. Then $A=B^2=BB=BTB$, so take $C=B$.

(c) $\Rightarrow$ (a)

Assume that $A=C^TC$, where $C$ is invertible.

$$x^TAx = x^TC^TCx = (Cx)^T(Cx) = Cx \cdot Cx = \|Cx\|^2 \geq 0$$

But the invertibility of $C$ implies that $Cx \neq 0$ if $x \neq 0$, so $x^TAx>0$ for $x \neq 0$. 
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

If a function $f(x, y)$ has first partial derivatives, then its relative maxima and minima, if any, occur at points where

$$f_x(x, y) = 0 \quad \text{and} \quad f_y(x, y) = 0$$

These are called **critical points** of $f$.

The specific behavior of $f$ at a critical point $(x_0, y_0)$ is determined by the sign of

$$D(x, y) = f(x, y) - f(x_0, y_0)$$

- If $D(x, y) > 0$ at points $(x, y)$ that are sufficiently close to, but different from, $(x_0, y_0)$, then $f(x_0, y_0) < f(x, y)$ at such points and $f$ is said to have a **relative minimum** at $(x_0, y_0)$. 

E-mail: hogijung@hanyang.ac.kr  
http://web.yonsei.ac.kr/hgjung
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

• If \( D(x,y) < 0 \) at points \((x,y)\) that are sufficiently close to, but different from, \((x_0,y_0)\), then \( f(x_0,y_0) > f(x,y) \) at such points and \( f \) is said to have a \textit{relative maximum} at \((x_0,y_0)\).

• If \( D(x,y) \) has both positive and negative values inside every circle centered at \((x_0,y_0)\), then there are points \((x,y)\) that are arbitrarily close to \((x_0,y_0)\) at which \( f(x_0,y_0) > f(x,y) \). In this case we say that \( f \) has a \textit{saddle point} at \((x_0,y_0)\).
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

**Theorem 8.5.1 (Second Derivative Test)** Suppose that \((x_0, y_0)\) is a critical point of \(f(x, y)\) and that \(f\) has continuous second-order partial derivatives in some circular region centered at \((x_0, y_0)\). Then:

(a) \(f\) has a relative minimum at \((x_0, y_0)\) if
\[
f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \text{and} \quad f_{xx}(x_0, y_0) > 0
\]

(b) \(f\) has a relative maximum at \((x_0, y_0)\) if
\[
f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \text{and} \quad f_{xx}(x_0, y_0) < 0
\]

(c) \(f\) has a saddle point at \((x_0, y_0)\) if
\[
f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) < 0
\]

(d) The test is inconclusive if
\[
f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) = 0
\]
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

To express this theorem in terms of quadratic forms, we consider the matrix

\[ H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{bmatrix} \]

which is called the Hessian or Hessian matrix of \( f \).

The Hessian is of interest because

\[
\det[H(x_0, y_0)] = \begin{vmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{vmatrix} = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2
\]

is the expression that appears in Theorem 8.5.1.
Relative Extrema of Functions of Two Variables

**Theorem 8.5.2 (Hessian Form of the Second Derivative Test)** Suppose that \((x_0, y_0)\) is a critical point of \(f(x, y)\) and that \(f\) has continuous second-order partial derivatives in some circular region centered at \((x_0, y_0)\). If \(H = H(x_0, y_0)\) is the Hessian of \(f\) at \((x_0, y_0)\), then:

(a) \(f\) has a relative minimum at \((x_0, y_0)\) if \(H\) is positive definite.

(b) \(f\) has a relative maximum at \((x_0, y_0)\) if \(H\) is negative definite.

(c) \(f\) has a saddle point at \((x_0, y_0)\) if \(H\) is indefinite.

(d) The test is inconclusive otherwise.

(a)

If \(H\) is positive definite, then Theorem 8.4.5 implies that the principal submatrices of \(H\) have positive determinants. Thus,

\[
\det[H] = \begin{vmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{vmatrix} = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0
\]

and

\[
\det[f_{xx}(x_0, y_0)] = f_{xx}(x_0, y_0) > 0
\]
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

Example 1

Find the critical points of the function

\[ f(x, y) = \frac{1}{3}x^3 + xy^2 - 8xy + 3 \]

and use the eigenvalues of the Hessian matrix at those points to determine which of them, if any, are relative maxima, relative minima, or saddle points.

\[
\begin{align*}
    f_x(x, y) &= x^2 + y^2 - 8y, & f_y(x, y) &= 2xy - 8x, & f_{xy}(x, y) &= 2y - 8 \\
    f_{xx}(x, y) &= 2x, & f_{yy}(x, y) &= 2x
\end{align*}
\]

Thus, the Hessian matrix is

\[
H(x, y) = \begin{bmatrix}
    f_{xx}(x, y) & f_{xy}(x, y) \\
    f_{xy}(x, y) & f_{yy}(x, y)
\end{bmatrix} = \begin{bmatrix}
    2x & 2y - 8 \\
    2y - 8 & 2x
\end{bmatrix}
\]
8.5. Application of Quadratic Forms to Optimization

Relative Extrema of Functions of Two Variables

Example 1

\( f_y(x, y) = 2xy - 8x = 2x(y - 4) = 0 \) \implies x=0 or y=4

\( f_x(x, y) = x^2 + y^2 - 8y = 0 \)

We have four critical points: \((0, 0), \ (0, 8), \ (4, 4), \ (-4, 4)\)

Evaluating the Hessian matrix at these points yields

\[
H(0, 0) = \begin{bmatrix} 0 & -8 \\ -8 & 0 \end{bmatrix}, \quad H(0, 8) = \begin{bmatrix} 0 & 8 \\ 8 & 0 \end{bmatrix}
\]

\[
H(4, 4) = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}, \quad H(-4, 4) = \begin{bmatrix} -8 & 0 \\ 0 & -8 \end{bmatrix}
\]

<table>
<thead>
<tr>
<th>정류점 ((x_0, y_0))</th>
<th>(\lambda_1)</th>
<th>(\lambda_2)</th>
<th>분류</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0)</td>
<td>8</td>
<td>-8</td>
<td>안정점</td>
</tr>
<tr>
<td>(0, 8)</td>
<td>8</td>
<td>-8</td>
<td>안정점</td>
</tr>
<tr>
<td>(4, 4)</td>
<td>8</td>
<td>8</td>
<td>극소값</td>
</tr>
<tr>
<td>(-4, 4)</td>
<td>-8</td>
<td>-8</td>
<td>극대값</td>
</tr>
</tbody>
</table>
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Geometrically, the problem of finding the maximum and minimum values of $x^T A x$ subject to the constraint $||x||=1$ amounts to finding the highest and lowest points on the intersections of the surface with the right circular cylinder determined by the circle.
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

**Theorem 8.5.3 (Constrained Extremum Theorem)** Let $A$ be a symmetric $n \times n$ matrix whose eigenvalues in order of decreasing size are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:

(a) There is a maximum value and a minimum value for $x^T A x$ on the unit sphere $||x|| = 1$.

(b) The maximum value is $\lambda_1$ (the largest eigenvalue), and this maximum occurs if $x$ is a unit eigenvector of $A$ corresponding to $\lambda_1$.

(c) The minimum value is $\lambda_n$ (the smallest eigenvalue), and this minimum occurs if $x$ is a unit eigenvector of $A$ corresponding to $\lambda_n$.

Since $A$ is symmetric, the principal axes theorem (Theorem 8.4.1) implies that there is an orthogonal change of variable $x = Py$ such that

$$x^T A x = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $A$.

Let us assume that $||x||=1$ and that the column vectors of $P$ (which are unit eigenvectors of $A$) have been ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Since \( P \) is an orthogonal matrix, multiplication by \( P \) is length preserving, so \( \|y\| = \|x\| = 1 \); that is,
\[
y_1^2 + y_2^2 + \cdots + y_n^2 = 1
\]

It follows from this equation and (7) that
\[
\lambda_n = \lambda_n(y_1^2 + y_2^2 + \cdots + y_n^2) \leq \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \leq \lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) = \lambda_1
\]
and hence from (6) that
\[
\lambda_n \leq x^T A x \leq \lambda_1
\]

This shows that all values of \( x^T A x \) for which \( \|x\| = 1 \) lie between the largest and smallest eigenvalues of \( A \).

Let \( x \) be a unit eigenvector corresponding to \( \lambda_1 \). Then
\[
x^T A x = x^T (\lambda_1 x) = \lambda_1 x^T x = \lambda_1 \|x\|^2 = \lambda_1
\]

If \( x \) is a unit eigenvector corresponding to \( \lambda_n \). Then
\[
x^T A x = x^T (\lambda_n x) = \lambda_n x^T x = \lambda_n \|x\|^2 = \lambda_n
\]
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Example 2

Find the maximum and minimum values of the quadratic form

\[ z = 5x^2 + 5y^2 + 4xy \]

subject to the constraint \( x^2 + y^2 = 1 \)

\[
\begin{align*}
  z &= 5x^2 + 5y^2 + 4xy = x^T A x = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
  \lambda_1 &= 7: \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 3: \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
  \lambda_1 &= 7: \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \lambda_2 = 3: \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}
\end{align*}
\]

Thus, the constrained extrema are

constrained maximum: \( z = 7 \) at \( (x, y) = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \)

constrained minimum: \( z = 3 \) at \( (x, y) = \left( -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \)
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Example 3

A rectangle is to be inscribed in the ellipse $4x^2 + 9y^2 = 36$, as shown in Figure 8.5.3. Use eigenvalue methods to find nonnegative values of $x$ and $y$ that produce the inscribed rectangle with maximum area.

The area $z$ of the inscribed rectangle is given by $z = 4xy$, so the problem is to maximize the quadratic form $z = 4xy$ subject to the constraint $4x^2 + 9y^2 = 36$.

\[
4x^2 + 9y^2 = 36 \implies \left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1 \implies x_1^2 + y_1^2 = 1
\]

\[
x = 3x_1
\]
\[
y = 2y_1
\]

\[
z = 4xy \implies z = 24x_1y_1
\]

\[
z = x^TAx = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} 0 & 12 \\ 12 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}
\]

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Example 3

The largest eigenvalue of $A$ is $\lambda = 12$ and that the only corresponding unit eigenvector with nonnegative entries is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Thus, the maximum area is $z = 12$, and this occurs when

$$x = 3x_1 = \frac{3}{\sqrt{2}} \quad \text{and} \quad y = 2y_1 = \frac{2}{\sqrt{2}}$$
8.5. Application of Quadratic Forms to Optimization

Constrained Extrema and Level Curves

Curves having equations of the form

\[ f(x, y) = k \]

are called the level curves of \( f \).

In particular, the level curves of a quadratic form \( x^T A x \) on \( R^2 \) have equations of the form

\[ x^T A x = k \quad (5) \]

so the maximum and minimum values of \( x^T A x \) subject to the constraint \( \|x\|=1 \) are the largest and smallest values of \( k \) for which the graph of (5) intersects the unit circle. Typically, such values of \( k \) produce level curves that just touch the unit circle. Moreover, the points at which these level curves just touch the circle produce the components of the vectors that maximize or minimize \( x^T A x \) subject to the constraint \( \|x\|=1 \).
8.5. Application of Quadratic Forms to Optimization

Constrained Extremum Problems

Example 4
In Example 2,
\[ z = 5x^2 + 5y^2 + 4xy \]
subject to the constraint \( x^2 + y^2 = 1 \)
constrained maximum: \( z = 7 \) at \( (x, y) = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \)
constrained minimum: \( z = 3 \) at \( (x, y) = \left( -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \)

Geometrically,

There can be no level curves
\[ 5x^2 + 5y^2 + 4xy = k \]
with \( k > 7 \) or \( k < 3 \) that intersects the unit circle.
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

We know from our work in Section 8.3 that symmetric matrices are orthogonally diagonalizable and are the only matrices with this property (Theorem 8.3.4).

The orthogonal diagonalizability of an $n \times n$ symmetric matrix $A$ means it can be factored as

$$A = PDPT$$

(1)

where $P$ is an $n \times n$ orthogonal matrix of eigenvectors of $A$, and $D$ is the diagonal matrix whose diagonal entries are the eigenvalues corresponding to the column vectors of $P$.

In this section we will call (1) an eigenvalue decomposition of $A$ (abbreviated EVD of $A$).
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

If an $n \times n$ matrix $A$ is not symmetric, then it does not have an eigenvalue decomposition, but it does have a Hessenberg decomposition

$$A = PHP^T$$

in which $P$ is an orthogonal matrix and $H$ is in upper Hessenberg form (Theorem 8.3.8).

Moreover, if $A$ has real eigenvalues, then it has a Schur decomposition

$$A = PSP^T$$

in which $P$ is an orthogonal matrix and $S$ is upper triangular (Theorem 8.3.7).
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

The eigenvalue, Hessenberg, and Schur decompositions are important in numerical algorithms not only because $H$ and $S$ have simpler forms than $A$, but also because the orthogonal matrices that appear in these factorizations do not magnify roundoff error.

Suppose that $\hat{x}$ is a column vector whose entries are known exactly and that 

$$x = \hat{x} + e$$

is the vector that results when roundoff error is present in the entries of $\hat{x}$.

If $P$ is an orthogonal matrix, then the length-preserving property of orthogonal transformations implies that 

$$\| Px - P\hat{x} \| = \| x - \hat{x} \| = \| e \|$$

which tells us that the error in approximating $P\hat{x}$ by $Px$ has the same magnitude as the error in approximating $\hat{x}$ by $x$. 
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

There are two main paths that one might follow in looking for other kinds of factorizations of a general square matrix $A$.

**Jordan canonical form**

$A = PJP^{-1}$

in which $P$ is invertible but not necessarily orthogonal. And, $J$ is either diagonal (Theorem 8.2.6) or a certain kind of block diagonal matrix.

**Singular Value Decomposition (SVD)**

$A = U\Sigma V^T$

in which $U$ and $V$ are orthogonal but not necessarily the same. And, $\Sigma$ is diagonal.
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

**Theorem 8.6.1** If $A$ is an $n \times n$ matrix of rank $k$, then $A$ can be factored as

$$A = U \Sigma V^T$$

where $U$ and $V$ are $n \times n$ orthogonal matrices and $\Sigma$ is an $n \times n$ diagonal matrix whose main diagonal has $k$ positive entries and $n - k$ zeros.

The matrix $A^T A$ is symmetric, so it has an eigenvalue decomposition

$$A^T A = V D V^T$$

where the column vectors of $V$ are unit eigenvectors of $A^T A$ and $D$ is the diagonal matrix whose diagonal entries are the corresponding eigenvalues of $A^T A$.

These eigenvalues are nonnegative, for if $\lambda$ is an eigenvalue of $A^T A$ and $x$ is a corresponding eigenvector, then Formula (12) of Section 3.2 implies that

$$\|Ax\|^2 = Ax \cdot Ax = x \cdot A^T A x = x \cdot \lambda x = \lambda (x \cdot x) = \lambda \|x\|^2$$

from which it follows that $\lambda \geq 0$. 
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

Since Theorems 7.5.8 and 8.2.3 imply that
\[ \text{rank}(A) = \text{rank}(A^T A) = \text{rank}(D) \]
and since \( A \) has rank \( k \), it follows that there are \( k \) positive entries and \( n-k \) zeros on the main diagonal of \( D \).

For convenience, suppose that the column vectors \( v_1, v_2, \ldots, v_n \) of \( V \) have been ordered so that the corresponding eigenvalues of \( A^T A \) are in nonincreasing order
\[ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \]
Thus,
\[ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0 \quad \text{and} \quad \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0 \]
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

Now consider the set of image vectors
\[ \{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\} \]
This is an orthogonal set, for if \(i \neq j\), then the orthogonality of \(\mathbf{v}_i\) and \(\mathbf{v}_j\) implies that
\[ A\mathbf{v}_i \cdot A\mathbf{v}_j = \mathbf{v}_i \cdot A^T A\mathbf{v}_j = \mathbf{v}_i \cdot \lambda_j \mathbf{v}_j = \lambda_j (\mathbf{v}_i \cdot \mathbf{v}_j) = 0 \quad (3) \]
Moreover,\[ \|A\mathbf{v}_i\|^2 = A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^T A\mathbf{v}_i = \mathbf{v}_i \cdot \lambda_i \mathbf{v}_i = \lambda_i (\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i \|\mathbf{v}_i\|^2 = \lambda_i \]
from which it follows that\[ \|A\mathbf{v}_i\| = \sqrt{\lambda_i} \quad (i = 1, 2, \ldots, n) \quad (4) \]
Since \(\lambda_i > 0\) for \(i = 1, 2, \ldots, k\), it follows from (3) and (4) that
\[ \{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_k\} \quad (5) \]
is an orthogonal set of \(k\) nonzero vectors in the column space of \(A\); and since we know that the column space of \(A\) has dimension \(k\) (since \(A\) has rank \(k\)), it follows that (5) is an orthogonal basis for the column space of \(A\).
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

If we normalize these vectors to obtain an orthonormal basis \( \{ \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \} \) for the column space, then Theorem 7.9.7 guarantees that we can extend this to an orthonormal basis

\[
\{ \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n \}
\]

for \( \mathbb{R}^n \).

Since the first \( k \) vectors in this set result from normalizing the vectors in (5), we have

\[
\mathbf{u}_j = \frac{A\mathbf{v}_j}{\| A\mathbf{v}_j \|} = \frac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j \quad (1 \leq j \leq k)
\]

which implies that

\[
A\mathbf{v}_1 = \sqrt{\lambda_1} \mathbf{u}_1, \quad A\mathbf{v}_2 = \sqrt{\lambda_2} \mathbf{u}_2, \ldots, \quad A\mathbf{v}_k = \sqrt{\lambda_k} \mathbf{u}_k
\]

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

Now let $U$ be the orthogonal matrix

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_k & u_{k+1} & \cdots & u_n \end{bmatrix}$$

and let $\Sigma$ be the diagonal matrix

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

It follows from (2) and (4) that $A v_j = 0$ for $j > k$, so

$$U \Sigma = \begin{bmatrix} \sqrt{\lambda_1} u_1 & \sqrt{\lambda_2} u_2 & \cdots & \sqrt{\lambda_k} u_k & 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} A v_1 & A v_2 & \cdots & A v_k & A v_{k+1} & \cdots & A v_n \end{bmatrix} = AV$$

which we can rewrite as $A = U \Sigma V^T$ using the orthogonality of $V$. 
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

It is important to keep in mind that the positive entries on the main diagonal of \( \Sigma \) are not eigenvalues of \( A \), but rather square roots of the nonzero eigenvalues of \( A^T A \). These numbers are called the singular values of \( A \) and are denoted by

\[
\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_k = \sqrt{\lambda_k}
\]

With this notation, the factorization obtained in the proof of Theorem 8.6.1 has the form

\[
A = U \Sigma V^T = [u_1 \ u_2 \ \cdots \ u_k \ u_{k+1} \ \cdots \ u_n]
\]

which is called the singular value decomposition of \( A \) (abbreviated SVD of \( A \)). The vectors \( u_1, u_2, \ldots, u_k \) are called left singular vectors of \( A \) and \( v_1, v_2, \ldots, v_k \) are called right singular vectors of \( A \).
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

**Theorem 8.6.2** *(Singular Value Decomposition of a Square Matrix)*  If $A$ is an $n \times n$ matrix of rank $k$, then $A$ has a singular value decomposition $A = U \Sigma V^T$ in which:

(a) $V = [v_1 \ v_2 \ \cdots \ v_n]$ orthogonally diagonalizes $A^T A$.

(b) The nonzero diagonal entries of $\Sigma$ are

$$\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of $A^T A$ corresponding to the column vectors of $V$.

(c) The column vectors of $V$ are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(d) $u_i = \frac{Av_i}{\|Av_i\|} = \frac{1}{\sigma_i} Av_i \quad (i = 1, 2, \ldots, k)$

(e) $\{u_1, u_2, \ldots, u_k\}$ is an orthonormal basis for $\text{col}(A)$.

(f) $\{u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_n\}$ is an extension of $\{u_1, u_2, \ldots, u_k\}$ to an orthonormal basis for $\mathbb{R}^n$. 
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

**Example 1**

Find the singular value decomposition of the matrix

\[ A = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \]

The first step is to find the eigenvalues of the matrix

\[ A^T A = \begin{bmatrix} \sqrt{3} & 0 \\ 2 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} 3 & 2\sqrt{3} \\ 2\sqrt{3} & 7 \end{bmatrix} \]

The characteristic polynomial of \( A^T A \) is

\[ \lambda^2 - 10\lambda + 9 = (\lambda - 9)(\lambda - 1) \]

so the eigenvalues of \( ATA \) are \( \lambda_1 = 9 \) and \( \lambda_2 = 1 \),

and the singular values of \( A \) are

\[ \sigma_1 = \sqrt{\lambda_1} = \sqrt{9} = 3, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1 \]
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

Example 1

Unit eigenvectors of $A^T A$ corresponding to the eigenvalues $\lambda_1 = 9$ and $\lambda_2 = 1$ are

$$v_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

respectively.

Thus,

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}, \quad u_2 = \frac{1}{\sigma_2} A v_2 = (1) \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

so

$$U = [u_1 \quad u_2] = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad V = [v_1 \quad v_2] = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$
8.6. Singular Value Decomposition

Singular Value Decomposition of Square Matrices

**Example 1**

It now follows that the singular value decomposition of \( A \) is

\[
\begin{bmatrix}
\sqrt{3} & 2 \\
0 & \sqrt{3}
\end{bmatrix} = \begin{bmatrix}
\frac{\sqrt{3}}{2} & -\frac{1}{2} \\
\frac{1}{2} & \frac{\sqrt{3}}{2}
\end{bmatrix} \begin{bmatrix}
3 & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
\frac{1}{2} & \frac{\sqrt{3}}{2} \\
-\frac{\sqrt{3}}{2} & \frac{1}{2}
\end{bmatrix}
\]

\( A = U \Sigma V^T \)
8.6. Singular Value Decomposition

**Singular Value Decomposition of Symmetric Matrices**

Suppose that \( A \) has rank \( k \) and that the nonzero eigenvalues of \( A \) are ordered so that
\[
|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k| > 0
\]

In the case where \( A \) is symmetric we have \( A^T A = A^2 \), so the eigenvalues of \( A^T A \) are the squares of the eigenvalues of \( A \).

Thus, the nonzero eigenvalues of \( A^T A \) in nonincreasing order are
\[
\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_k^2 > 0
\]
and the singular values of \( A \) in nonincreasing order are
\[
\sigma_1 = \sqrt{\lambda_1^2} = |\lambda_1|, \quad \sigma_2 = \sqrt{\lambda_2^2} = |\lambda_2|, \ldots, \quad \sigma_k = \sqrt{\lambda_k^2} = |\lambda_k|
\]

This shows that the *singular values of a symmetric matrix \( A \) are the absolute values of the nonzero eigenvalues of \( A \); and it also shows that if \( A \) is a symmetric matrix with nonnegative eigenvalues, then the singular values of \( A \) are the same as its nonzero eigenvalues.*
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

If $A$ is an $m \times n$ matrix, then $A^T A$ is an $n \times n$ symmetric matrix and hence has an eigenvalue decomposition, just as in the case where $A$ is square.

Except for appropriate size adjustments to account for the possibility that $n>m$ or $n<m$, the proof of Theorem 8.6.1 carries over without change and yields the following generalization of Theorem 8.6.2.
Singular Value Decomposition of Nonsquare Matrices

Theorem 8.6.4 (Singular Value Decomposition of a General Matrix) If $A$ is an $m \times n$ matrix of rank $k$, then $A$ can be factored as

$$A = U \Sigma V^T = [u_1 \ u_2 \ \cdots \ u_k \mid u_{k+1} \ \cdots \ u_m]$$

$$= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \\ 0_{(m-k) \times k} & 0_{(m-k) \times (n-k)} \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \\ v_{k+1}^T \\ \vdots \\ v_n^T \end{bmatrix}$$

where in which $U$, $\Sigma$, and $V$ have sizes $m \times m$, $m \times n$, and $n \times n$, respectively, and in which:

(a) $V = [v_1 \ v_2 \ \cdots \ v_n]$ orthogonally diagonalizes $A^T A$.

(b) The nonzero diagonal entries of $\Sigma$ are $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}$, \ldots, $\sigma_k = \sqrt{\lambda_k}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of $A^T A$ corresponding to the column vectors of $V$.

(c) The column vectors of $V$ are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(d) $u_i = \frac{A v_i}{\| A v_i \|} = \frac{1}{\sigma_i} A v_i \quad (i = 1, 2, \ldots, k)$

(e) $\{u_1, u_2, \ldots, u_k\}$ is an orthonormal basis for col$(A)$.

(f) $\{u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_m\}$ is an extension of $\{u_1, u_2, \ldots, u_k\}$ to an orthonormal basis for $R^n$. 
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

\[ A = U \Sigma V^T = [u_1 \ u_2 \ \cdots \ u_k \ | \ u_{k+1} \ \cdots \ u_m] \]

**singular values** of \( A \)

**left singular vectors** of \( A \)

**right singular vectors** of \( A \)

\[
\begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_k \\
\end{bmatrix}
\begin{bmatrix}
\theta_{(m-k)\times k} & 0_{k\times(n-k)} \\
\vdots & \vdots \\
0_{k\times(n-k)} & \theta_{(m-k)\times(n-k)} \\
\end{bmatrix}
\begin{bmatrix}
v_1^T \\
v_2^T \\
v_k^T \\
v_{k+1}^T \\
v_n^T \\
\end{bmatrix}
\]
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

Example 4

Find the singular value decomposition of the matrix

\[ A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \]

\[ A^T A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \]

\[ \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1) \]

\[ \sigma_1 = \sqrt{\lambda_1} = \sqrt{3}, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1 \]

\[ v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \quad v_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \]
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

Example 4

$$u_1 = \frac{1}{\sigma_1} Av_1 = \frac{\sqrt{3}}{3} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}, \quad u_2 = \frac{1}{\sigma_2} Av_2 = (1) \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{6}}{2} \\ -\frac{\sqrt{6}}{2} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\sqrt{6} u_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad \sqrt{2} u_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

A unit vector $u_3$ that is orthogonal to $u_1$ and $u_2$ must be a solution of the homogeneous linear system

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad u_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

Example 4

Thus, the singular value decomposition of $A$ is

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

$$
A = U \Sigma V^T
$$
8.6. Singular Value Decomposition

Singular Value Decomposition of Nonsquare Matrices

Theorem 8.6.5  If $A$ is an $m \times n$ matrix with rank $k$, and if $A = U \Sigma V^T$ is the singular value decomposition given in Formula (12), then:

(a) $\{u_1, u_2, \ldots, u_k\}$ is an orthonormal basis for $\text{col}(A)$.
(b) $\{u_{k+1}, \ldots, u_m\}$ is an orthonormal basis for $\text{col}(A)^\perp = \text{null}(A^T)$.
(c) $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal basis for $\text{row}(A)$.
(d) $\{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $\text{row}(A)^\perp = \text{null}(A)$. 
8.6. Singular Value Decomposition

Reduced Singular Value Decomposition

The products that involve zero blocks as factors drop out, leaving

\[ A = [u_1 \quad u_2 \quad \cdots \quad u_k] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix} [v_1^T \quad v_2^T \quad \cdots \quad v_k^T] \]

which is called a \textit{reduced singular value decomposition} of \( A \)

\[ A = U_1 \Sigma_1 V_1^T \]

\( U_1 \): \( m \times k \), \( \Sigma_1 \): \( k \times k \), and \( V_1 \): \( k \times n \)

\[ A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T \]

which is called a \textit{reduced singular value expansion} of \( A \)
8.6. Singular Value Decomposition

Data Compression and Image Processing

If the matrix $A$ has size $m \times n$, then one might store each of its $mn$ entries individually.

An alternative procedure is to compute the reduced singular value decomposition

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T \quad (17)$$

in which $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, and store the $\sigma$’s, the $u$’s and the $v$’s.

When needed, the matrix $A$ can be reconstructed from (17). Since each $u_j$ has $m$ entries and each $v_j$ has $n$ entries, this method requires storage space for

$$km + kn + k = k(m + n + 1)$$

numbers.
8.6. Singular Value Decomposition

Data Compression and Image Processing

\[ A_r = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \]

It is called the rank \( r \) approximation of \( A \). This matrix requires storage space for only

\[ rm + rn + r = r(m + n + 1) \]

numbers, compared to \( mn \) numbers required for entry-by-entry storage of \( A \).

For example, the rank 100 approximation of a \( 1000 \times 1000 \) matrix \( A \) requires storage for only

\[ 100(1000 + 1000 + 1) = 200,100 \]

numbers, compared to the 1,000,000 numbers required for entry-by-entry storage of \( A \) – a compression of almost 80%.

rank 4  rank 10  rank 20  rank 50  rank 128
8.6. Singular Value Decomposition

Singular Value Decomposition from The Transformation Point of View

If $A$ is an $m \times n$ matrix and $T_A : R^n \rightarrow R^m$ is multiplication by $A$, then the matrix

$$
\Sigma = \begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_k \\
\end{bmatrix}
\begin{bmatrix}
0_{k \times (n-k)} \\
\vdots \\
0_{(m-k) \times k} \\
0_{(m-k) \times (n-k)}
\end{bmatrix}
$$

in (12) is the matrix for $T_A$ with respect to the bases $\{v_1, v_2, \ldots, v_n\}$ and $\{u_1, u_2, \ldots, u_m\}$ for $R^n$ and $R^m$, respectively.
8.6. Singular Value Decomposition

Singular Value Decomposition from The Transformation Point of View

Thus, when vectors are expressed in terms of these bases, we see that the effect of multiplying a vector by $A$ is to scale the first $k$ coordinates of the vector by the factor $\sigma_1, \sigma_2, \ldots, \sigma_k$, map the rest of the coordinates to zero, and possibly to discard coordinates or append zeros, if needed, to account for a decrease or increase in dimension.

This idea is illustrated in Figure 8.6.4 for a $2 \times 3$ matrix $A$ of rank 2.
8.6. Singular Value Decomposition

Singular Value Decomposition from The Transformation Point of View

Since \( \text{row}(A) \perp \text{null}(A) \), it follows from Theorem 7.7.4 that every vector \( x \) in \( \mathbb{R}^n \) can be expressed uniquely as
\[
X = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)}
\]
where \( \mathbf{x}_{\text{row}(A)} \) is the orthogonal projection of \( x \) on the row space of \( A \) and \( \mathbf{x}_{\text{null}(A)} \) is its orthogonal projection on the null space of \( A \).

Since \( A \mathbf{x}_{\text{null}(A)} = 0 \), it follows that
\[
T_A(x) = Ax = A\mathbf{x}_{\text{row}(A)} + A\mathbf{x}_{\text{null}(A)} = A\mathbf{x}_{\text{row}(A)}
\]
8.6. Singular Value Decomposition

Singular Value Decomposition from The Transformation Point of View

This tells us three things:

1. The image of any vector in $\mathbb{R}^n$ under multiplication by $A$ is the same as the image of the orthogonal projection of that vector on $\text{row}(A)$.

2. The range of the transformation $T_A$, namely $\text{col}(A)$, is the image of $\text{row}(A)$.

3. $T_A$ maps distinct vectors in $\text{row}(A)$ into distinct vectors in $\mathbb{R}^m$. Thus, even though $T_A$ may not be one-to-one when considered as a transformation with domain $\mathbb{R}^n$, it is one-to-one if its domain is restricted to $\text{row}(A)$.

REMARK

“hiding” inside of every nonzero matrix transformation $T_A$ there is a one-to-one matrix transformation that maps the row space of $A$ onto the column space of $A$. Moreover, that hidden transformation is represented by the reduced singular value decomposition of $A$ with respect to appropriate bases.
8.7. The Pseudoinverse

The Pseudoinverse

If $A$ is an invertible $n \times n$ matrix with reduced singular value decomposition

$$A = U_1 \Sigma_1 V_1^T$$

then $U_1$, $\Sigma_1$, $V_1$ are all $n \times n$ invertible matrices, so the orthogonality of $U_1$ and $V_1$ implies that

$$A^{-1} = V_1 \Sigma_1^{-1} U_1^T$$  \hspace{1cm} (1)

If $A$ is not square or if it is square but not invertible, then this formula does not apply.

As the matrix $\Sigma_1$ is always invertible, the right side of (1) is defined for every matrix $A$.

If $A$ is nonzero $m \times n$ matrix, then we call the $n \times m$ matrix

$$A^+ = V_1 \Sigma_1^{-1} U_1^T$$  \hspace{1cm} (2)

the *pseudoinverse* of $A$. 
8.7. The Pseudoinverse

The Pseudoinverse

**Theorem 8.7.1** If $A$ is an $m \times n$ matrix with full column rank, then

$$A^+ = (A^TA)^{-1}A^T$$

Since $A$ has full column rank, the matrix $A^TA$ is invertible (Theorem 7.5.10) and $V_1$ is an $n \times n$ orthogonal matrix. Thus,

$$(A^TA)^{-1} = V_1 \Sigma_1^{-2} V_1^T$$

from which it follows that

$$(A^TA)^{-1}A^T = (V_1 \Sigma_1^{-2} V_1^T)(U_1 \Sigma_1 V_1^T)^T = (V_1 \Sigma_1^{-2} V_1^T)(V_1 \Sigma_1 U_1^T) = V_1 \Sigma_1^{-1} U_1^T = A^+$$
8.7. The Pseudoinverse

The Pseudoinverse

Example 1

Find the pseudoinverse of the matrix

\[
A = \begin{bmatrix}
1 & 1 \\
0 & 1 \\
1 & 0 \\
\end{bmatrix}
\]

using the reduced singular value decomposition that was obtained in Example 5 of Section 8.6.

\[
A^+ = [v_1 \ v_2] \begin{bmatrix}
\frac{1}{\sigma_1} & 0 \\
0 & \frac{1}{\sigma_2} \\
\end{bmatrix} [u_1^T \ u_2^T] = [\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \frac{1}{\sqrt{3}} & 0 \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \frac{0}{0} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \frac{3}{3} & -\frac{1}{3}
\end{bmatrix}
\]
The Pseudoinverse

**Example 2**

\[
A = \begin{bmatrix}
1 & 1 \\
0 & 1 \\
1 & 0
\end{bmatrix}
\]

A has full column rank so its pseudoinverse can also be computed from Formula (3).

\[
A^T A = \begin{bmatrix}
1 & 0 & 1 \\
1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
0 & 1 \\
1 & 0
\end{bmatrix} = \begin{bmatrix}
2 & 1 \\
1 & 2
\end{bmatrix}
\]

\[
A^+ = (A^T A)^{-1} A^T = \begin{bmatrix}
\frac{2}{3} & -\frac{1}{3} \\
-\frac{1}{3} & \frac{2}{3}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}
\end{bmatrix}
\]
### 8.7. The Pseudoinverse

#### Properties of The Pseudoinverse

**Theorem 8.7.2**  If $A^+$ is the pseudoinverse of an $m \times n$ matrix $A$, then:

- (a) $AA^+A = A$
- (b) $A^+AA^+ = A^+$
- (c) $(AA^+)^T = AA^+$
- (d) $(A^+A)^T = A^+A$
- (e) $(A^T)^+ = (A^+)^T$
- (f) $A^{++} = A$
8.7. The Pseudoinverse

Properties of The Pseudoinverse

**Theorem 8.7.3** If $A^+ = V_1 \Sigma_1^{-1} U_1^T$ is the pseudoinverse of an $m \times n$ matrix $A$ of rank $k$, and if the column vectors of $U_1$ and $V_1$ are $u_1, u_2, \ldots, u_k$ and $v_1, v_2, \ldots, v_k$, respectively, then:

(a) $A^+ y$ is in row($A$) for every vector $y$ in $R^m$.

(b) $A^+ u_i = \frac{1}{\sigma_i} v_i$ \hspace{1cm} (i = 1, 2, \ldots, k)

(c) $A^+ y = 0$ for every vector $y$ in null($A^T$).

(d) $AA^+$ is the orthogonal projection of $R^m$ onto col($A$).

(e) $A^+ A$ is the orthogonal projection of $R^n$ onto row($A$).

(a) If $y$ is a vector in $R^m$, then it follows from (2) that

$$A^+ y = V_1 \Sigma_1^{-1} U_1^T y = V_1 (\Sigma_1^{-1} U_1^T y)$$

so $A^+ y$ must be a linear combination of the column vectors of $V_1$. Since Theorem 8.6.5 states that these vectors are in row($A$), it follows that $A^+ y$ is in row($A$).
8.7. The Pseudoinverse

Properties of The Pseudoinverse

(b) Multiplying $A^+$ on the right by $U_1$ yields

$$A^+ U_1 = V_1 \Sigma_1^{-1} U_1^T U_1 = V_1 \Sigma_1^{-1}$$

The result now follows by comparing corresponding column vectors on the two sides of this equation.

(c) If $y$ is a vector in $\text{null}(A^T)$, then $y$ is orthogonal to each vector in $\text{col}(A)$, and, in particular, it is orthogonal to each column vector of $U_1=[u_1, u_2, \ldots, u_k]$.

This implies that $U_1^T y = 0$, and hence that

$$A^+ y = V_1 \Sigma_1^{-1} U_1^T y = (V_1 \Sigma_1^{-1}) U_1^T y = 0$$
8.7. The Pseudoinverse

Properties of The Pseudoinverse

Example 3

Use the pseudoinverse of

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

to find the standard matrix for the orthogonal projection of $\mathbb{R}^3$ onto the column space of $A$.

The pseudoinverse of $A$ was computed in Example 2.

Using that result we see that the orthogonal projection of $\mathbb{R}^3$ onto $\text{col}(A)$ is

$$AA^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$
8.7. The Pseudoinverse

Pseudoinverse And Least Squares

Linear system $Ax=b$

Least square solutions: the exact solution of the normal equation $A^TAx=ATb$

- If full column rank, there is a unique least square solution $x=(A^TA)^{-1}ATb=A^+b$
- If not, there are infinitely many solutions

We know that **among these least squares solutions there is a unique least square solution in the row space of $A$** (Theorem 7.8.3), and we also know that **it is the least squares solution of minimum norm**.
8.7. The Pseudoinverse

Pseudoinverse And Least Squares

**Theorem 8.7.4** If $A$ is an $m \times n$ matrix, and $b$ is any vector in $\mathbb{R}^m$, then 

$$x = A^+b$$

is the least squares solution of $Ax = b$ that has minimum norm.

(1) Let $A = U_1 \Sigma_1 V_1^T$ be a reduced singular value decomposition of $A$, so 

$$A^+b = V_1 \Sigma_1^{-1} U_1^T b$$

Thus,

$$(A^T A)^+ b = V_1 \Sigma_1^2 V_1^T V_1 \Sigma_1^{-1} U_1^T b = V_1 \Sigma_1^2 \Sigma_1^{-1} U_1^T b = V_1 \Sigma_1 U_1^T b = A^T b$$

which shows that $x = A^+b$ satisfies the normal equation and hence is a least squares solution.

(2) $x = A^+b$ lies in the row space of $A$ by part (a) of Theorem 8.7.3.

Thus, it is the least squares solution of minimum norm (Theorem 7.8.3).
8.7. The Pseudoinverse

Pseudoinverse And Least Squares

$x^+$ is the least squares solution of minimum norm and is an exact solution of the linear system; that is

$$Ax^+ = b_{\text{col}(A)}$$

$$A^+Ax^+ = A^+b_{\text{col}(A)} \rightarrow x^+ = A^+b_{\text{col}(A)}$$
8.7. The Pseudoinverse

Condition Number And Numerical Considerations

\[ Ax = b \]

If \( A \) is “nearly singular”, the linear system is said to be \textit{ill conditioned}.

A good measure of how roundoff error will affect the accuracy of a computed solution is given by the ratio of the largest singular value of \( A \) to the smallest singular values of \( A \).

It is called \textit{condition number} of \( A \), is denoted by

\[
\text{cond}(A) = \frac{\sigma_1}{\sigma_k}
\]

The larger the condition number, the more sensitive the system to small roundoff errors.
8.8. Complex Eigenvalues and EigenVectors

Vectors in $C^n$

**Definition 8.8.1** If $n$ is a positive integer, then a *complex $n$-tuple* is a sequence of $n$ complex numbers $(v_1, v_2, \ldots, v_n)$. The set of all complex $n$-tuples is called *complex $n$-space* and is denoted by $C^n$.

**modulus (or absolute value)**

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

**polar form**

$$z = |z| \cos \phi + i |z| \sin \phi$$

**complex conjugate**

$$\bar{v} = (\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n) = (a_1 - b_1 i, a_2 - b_2 i, \ldots, a_n - b_n i)$$

The standard operations on real matrices carry over to complex matrices without change, and all of the familiar properties of matrices continue to hold.
8.8. Complex Eigenvalues and EigenVectors

Algebraic Properties of The Complex Conjugate

Theorem 8.8.2  If \( \mathbf{u} \) and \( \mathbf{v} \) are vectors in \( \mathbb{C}^n \), and if \( k \) is a scalar, then:

(a) \( \overline{\mathbf{u}} = \mathbf{u} \)
(b) \( \overline{k\mathbf{u}} = k\overline{\mathbf{u}} \)
(c) \( \overline{\mathbf{u} + \mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}} \)
(d) \( \overline{\mathbf{u} - \mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}} \)

Theorem 8.8.3  If \( A \) is an \( m \times k \) complex matrix and \( B \) is a \( k \times n \) complex matrix, then:

(a) \( \overline{A} = A \)
(b) \( (A^T) = (\overline{A})^T \)
(c) \( \overline{AB} = \overline{A} \overline{B} \)
The analogous formulas in $C^n$ are

$$u \cdot v = u^T \bar{v} = \bar{v}^T u$$
8.8. Complex Eigenvalues and EigenVectors

The Complex Euclidean Inner Product

Example 2
\[ \mathbf{u} = (1 + i, i, 3 - i) \]
\[ \mathbf{v} = (1 + i, 2, 4i) \]

\[ \mathbf{u} \cdot \mathbf{v} = (1 + i)(1 + i) + i(2) + (3 - i)(4i) = (1 + i)(1 - i) + 2i + (3 - i)(-4i) = -2 - 10i \]
\[ \mathbf{v} \cdot \mathbf{u} = (1 + i)(1 + i) + 2(i) + (4i)(3 - i) = (1 + i)(1 - i) - 2i + 4i(3 + i) = -2 + 10i \]
\[ \|\mathbf{u}\| = \sqrt{|1 + i|^2 + |i|^2 + |3 - i|^2} = \sqrt{2 + 1 + 10} = \sqrt{13} \]
\[ \|\mathbf{v}\| = \sqrt{|1 + i|^2 + |2|^2 + |4i|^2} = \sqrt{2 + 4 + 16} = \sqrt{22} \]
8.8. Complex Eigenvalues and EigenVectors

The Complex Euclidean Inner Product

**Theorem 8.8.5** If \( \mathbf{u} \), \( \mathbf{v} \), and \( \mathbf{w} \) are vectors in \( \mathbb{C}^n \), and if \( k \) is a scalar, then the complex Euclidean inner product has the following properties:

(a) \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \) \hspace{1cm} \text{[Antisymmetry property]}

(b) \( \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \) \hspace{1cm} \text{[Distributive property]}

(c) \( k(\mathbf{u} \cdot \mathbf{v}) = (ku) \cdot \mathbf{v} \) \hspace{1cm} \text{[Homogeneity property]}

(d) \( \mathbf{v} \cdot \mathbf{v} \geq 0 \) and \( \mathbf{v} \cdot \mathbf{v} = 0 \) if and only if \( \mathbf{v} = \mathbf{0} \) \hspace{1cm} \text{[Positivity property]}

(c) \( k(\mathbf{u} \cdot \mathbf{v}) = k(\mathbf{v} \cdot \mathbf{u}) = \overline{k} (\overline{\mathbf{v}} \cdot \mathbf{u}) = \overline{k} (\mathbf{v} \cdot \mathbf{u}) = (\overline{k \mathbf{v}}) \cdot \mathbf{u} = \mathbf{u} \cdot (\overline{k \mathbf{v}}) \)

\[ \mathbf{u} \cdot k \mathbf{v} = \overline{k} (\mathbf{u} \cdot \mathbf{v}) \]
8.8. Complex Eigenvalues and EigenVectors

Vector Space Concepts in $C^n$

Except for the use of complex scalars, notions of linear combination, linear independence, subspace, spanning, basis, and dimension carry over without change to $C^n$, as do most of the theorems we have given in this text about them.
8.8. Complex Eigenvalues and EigenVectors

Complex Eigenvalues of Real Matrices Acting on Vectors in $\mathbb{C}^n$

*complex eigenvalue, complex eigenvector, eigenspace*

**Theorem 8.8.6**  If $\lambda$ is an eigenvalue of a real $n \times n$ matrix $A$, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of $A$, and $\overline{\mathbf{x}}$ is a corresponding eigenvector.

\[ A\mathbf{x} = \lambda\mathbf{x} \quad \Rightarrow \quad \overline{A}\mathbf{x} = \overline{\lambda}\mathbf{x} = \overline{\lambda}\overline{\mathbf{x}} \quad \Rightarrow \quad \overline{A\mathbf{x}} = A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}} \]

\[ A = \overline{A} \quad \text{since} \ A \ \text{has real entries} \]

\[ \overline{A\mathbf{x}} = A\overline{\mathbf{x}} = A\overline{\mathbf{x}} \]
8.8. Complex Eigenvalues and EigenVectors

Complex Eigenvalues of Real Matrices Acting on Vectors in $\mathbb{C}^n$

Example 3

Find the eigenvalues and bases for the eigenspaces of

$$A = \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix}$$

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

With $\lambda = i$,

$$\begin{bmatrix} i + 2 & 1 \\ -5 & i - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} i + 2 & 1 & 0 \\ -5 & i - 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{2}{5} - \frac{1}{5}i \\ 0 \end{bmatrix} \Rightarrow x_1 = \left( -\frac{2}{5} + \frac{1}{5}i \right) t$$

$$x_2 = t$$

With $\lambda = -i$,

$$\overline{x} = \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix}$$

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.8. Complex Eigenvalues and EigenVectors

A Proof That Real Symmetric Matrices Have Real Eigenvalues

**Theorem 8.8.7** If $A$ is a real symmetric matrix, then $A$ has real eigenvalues.

Suppose that $\lambda$ is an eigenvalue of $A$ and $x$ is a corresponding eigenvector.

$$Ax = \lambda x$$

where $x \neq 0$.

$$\overline{x}^T Ax = \overline{x}^T (\lambda x) = \lambda (\overline{x}^T x) = \lambda (x \cdot x) = \lambda \|x\|^2 \Rightarrow \lambda = \frac{\overline{x}^T Ax}{\|x\|^2}$$

$$\overline{x}^T Ax = \overline{x}^T \overline{Ax} = x^T \overline{Ax} = (\overline{Ax})^T x = (\overline{A\overline{x}})^T x = (A\overline{x})^T x = \overline{x}^T A^T x = \overline{x}^T Ax$$

which shows the numerator is real.

Since the denominator is real, $\lambda$ is real.
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

Theorem 8.8.8  The eigenvalues of the real matrix

\[
C = \begin{bmatrix}
  a & -b \\
  b & a
\end{bmatrix}
\]

are \( \lambda = a \pm bi \). If \( a \) and \( b \) are not both zero, then this matrix can be factored as

\[
\begin{bmatrix}
  a & -b \\
  b & a
\end{bmatrix} = \begin{bmatrix}
  |\lambda| & 0 \\
  0 & |\lambda|
\end{bmatrix} \begin{bmatrix}
  \cos \phi & -\sin \phi \\
  \sin \phi & \cos \phi
\end{bmatrix}
\]

where \( \phi \) is the angle from the positive \( x \)-axis to the ray from the origin through the point \((a, b)\) (Figure 8.8.2).

Geometrically, this theorem states that multiplication by a matrix of form (19) can be viewed as a rotation through the angle \( \phi \) followed by a scaling with factor \(|\lambda|\).
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

**Theorem 8.8.9**  Let $A$ be a real $2 \times 2$ matrix with complex eigenvalues $\lambda = a \pm bi$ (where $b \neq 0$). If $x$ is an eigenvector of $A$ corresponding to $\lambda = a - bi$, then the matrix $P = [\Re(x) \ \Im(x)]$ is invertible and

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1} \tag{21}$$

This theorem shows that every real $2 \times 2$ matrix with complex eigenvalues is similar to a matrix form (19).
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

\[ A = P S R_\phi P^{-1} = P \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} P^{-1} \]  \hspace{1cm} (22)

If we now view \( P \) as the transition matrix from the basis \( B=\{\text{Re}(x), \text{Im}(x)\} \) to the standard basis, then (22) tells us that computing a product \( Ax_0 \) can be broken down into a three-step process:

1. Map \( x_0 \) from standard coordinates into \( B \)-coordinates by forming the product \( P^{-1}x_0 \).
2. Rotate and scale the vector \( P^{-1}x_0 \) by forming the product \( SR_\phi P^{-1}x_0 \).
3. Map the rotated and scaled vector back to standard coordinates to obtain

\[ Ax_0 = P S R_\phi P^{-1}x_0 \]
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

Example 5

\[ A = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \]

A has eigenvalues \( \lambda = \frac{4}{5} \pm \frac{3}{5}i \) and that corresponding eigenvectors are

\[ \lambda_1 = \frac{4}{5} - \frac{3}{5}i: \quad \mathbf{v}_1 = \left( \frac{1}{2} + i, 1 \right), \quad \lambda_2 = \frac{4}{5} + \frac{3}{5}i: \quad \mathbf{v}_2 = \left( \frac{1}{2} - i, 1 \right) \]

If we take \( \lambda = \lambda_1 = \frac{4}{5} - \frac{3}{5}i \) and \( \mathbf{x} = \mathbf{v}_1 = \left( \frac{1}{2} + i, 1 \right) \) in (21)

and use the fact that \(|\lambda|=1\), then

\[ A = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}, \quad P = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix} \]

\[ \tan \phi = \frac{\sin \phi}{\cos \phi} = \frac{3/5}{4/5} = \frac{3}{4} \quad (\phi = \tan^{-1} \frac{3}{4} \approx 36.9^\circ) \]
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

Example 5

The matrix $P$ in (23) is the transition matrix from the basis

$$B = \{\text{Re}(x), \text{Im}(x)\} = \left\{ \left( \frac{1}{2}, 1 \right), (1, 0) \right\}$$

to the standard basis, and $P^{-1}$ is the transition matrix from the standard basis to the basis $B$.

In $B$-coordinates each successive multiplication by $A$ causes the point $P^{-1}x_0$ to advance through an angle, thereby tracing a circular orbit about the origin.

However, the basis $B$ is skewed (not orthogonal), so when the points on the circular orbit are transformed back to standard coordinates, the effect is to distort the circular orbit into the elliptical orbit.
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

Example 5

\[
\begin{bmatrix}
\frac{1}{2} & \frac{3}{4} \\
-\frac{3}{5} & \frac{11}{10}
\end{bmatrix}
\begin{bmatrix}
1 \\
1
\end{bmatrix}
= 
\begin{bmatrix}
\frac{1}{2} & 1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{4}{5} & -\frac{3}{5} \\
\frac{3}{5} & \frac{4}{5}
\end{bmatrix}
\begin{bmatrix}
0 & 1 \\
1 & -\frac{1}{2}
\end{bmatrix}
\begin{bmatrix}
1 \\
1
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
\frac{1}{2} & 1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{4}{5} & -\frac{3}{5} \\
\frac{3}{5} & \frac{4}{5}
\end{bmatrix}
\begin{bmatrix}
1 \\
\frac{1}{2}
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
\frac{1}{2} & 1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{5}{4} \\
\frac{1}{2}
\end{bmatrix}
\]

\[x_0 \text{ is mapped to } B\text{-coordinates}\]

\[\text{[The point } (1,1/2) \text{ is rotated through the angle } \phi]\]

\[\text{[The point } (1/2,1) \text{ is mapped to standard coordinates]}\]
8.8. Complex Eigenvalues and EigenVectors

A Geometric Interpretation of Complex Eigenvalues of Real Matrices

\[ A^n x_0 = (P R_\phi P^{-1})^n x_0 = P R_\phi^n P^{-1} x_0 \]

Section 6.1, Power Sequences
Hermitian and Unitary Matrices

**Definition 8.9.1** If $A$ is a complex matrix, then the **conjugate transpose** of $A$, denoted by $A^*$, is defined by

$$A^* = \overline{A}^T$$

(REMARK

Part (b) of Theorem 8.8.3 \((\overline{A^T}) = (\overline{A})^T\)

The order in which the transpose and conjugate operations are performed in computing $A^* = \overline{A}^T$ does not matter.

In the case where $A$ has real entries we have $A^* = A^T$,

so, $A^*$ is the same as $A^T$ for real matrices.)
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Example 1

Find the conjugate transpose $A^*$ of the matrix

\[
A = \begin{bmatrix}
1 + i & -i & 0 \\
2 & 3 - 2i & i
\end{bmatrix}
\]

\[
\overline{A} = \begin{bmatrix}
1 - i & i & 0 \\
2 & 3 - 2i & i
\end{bmatrix}
\]

and hence

\[
A^* = \begin{bmatrix}
1 - i & 2 \\
i & 3 + 2i \\
0 & -i
\end{bmatrix}
\]
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Theorem 8.9.2 If \( k \) is a complex scalar, and if \( A, B, \) and \( C \) are complex matrices whose sizes are such that the stated operations can be performed, then:

\[
\begin{align*}
(a) \quad (A^*)^* &= A \\
(b) \quad (A + B)^* &= A^* + B^* \\
(c) \quad (A - B)^* &= A^* - B^* \\
(d) \quad (kA)^* &= \overline{k} A^* \\
(e) \quad (AB)^* &= B^* A^*
\end{align*}
\]

\[
u \cdot v = \overline{v^T u} \quad \rightarrow \quad v^* u
\]

Formula (9) of Section 8.8
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

**Definition 8.9.3** A square complex matrix $A$ is said to be *unitary* if

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

(3)

and is said to be *Hermitian* if

$$A^* = A$$

(4)

The *unitary* matrices are complex generalizations of the real *orthogonal* matrices and *Hermitian* matrices are complex generalization of the real *symmetric* matrices.
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Example 2

\[
A = \begin{bmatrix}
1 & i & 1 + i \\
-i & -5 & 2 - i \\
1 - i & 2 + i & 3
\end{bmatrix}
\]

\[
\overline{A} = \begin{bmatrix}
1 & -i & 1 - i \\
i & -5 & 2 + i \\
1 + i & 2 - i & 3
\end{bmatrix}
\]

\[
A^* = \overline{A}^T = \begin{bmatrix}
1 & i & 1 + i \\
-i & -5 & 2 - i \\
1 - i & 2 + i & 3
\end{bmatrix} = A
\]

Thus, \( A \) is Hermitian.
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Theorem 8.9.4  The eigenvalues of a Hermitian matrix are real numbers.

The fact that real symmetric matrices have real eigenvalues is a special case of the more general result about Hermitian matrices.

Theorem 8.9.5  If $$A$$ is a Hermitian matrix, then eigenvectors from different eigenspaces are orthogonal.

Let $$v_1$$ and $$v_2$$ be eigenvectors corresponding to distinct eigenvalues $$\lambda_1$$ and $$\lambda_2$$.

$$A=A^*, \lambda_1 = \overline{\lambda_1}, \lambda_2 = \overline{\lambda_2}$$

$$\lambda_1 (v_2 \cdot v_1) = (\lambda_1 v_1)^* v_2 = (Av_1)^* v_2 = (v_1^* A^*) v_2$$

$$= (v_1^* A) v_2 = v_1^* (A v_2)$$

$$= v_1^* (\lambda_2 v_2) = \lambda_2 (v_1^* v_2) = \lambda_2 (v_2 \cdot v_1)$$

This implies that $$(\lambda_1 - \lambda_2) v_2 \cdot v_1 = 0$$ and hence that $$v_2 \cdot v_1 = 0$$ (since $$\lambda_1 \neq \lambda_2$$).
Hermitian and Unitary Matrices

Example 3

Confirm that the Hermitian matrix

\[
A = \begin{bmatrix}
2 & 1 + i \\
1 - i & 3
\end{bmatrix}
\]

has real eigenvalues and that eigenvectors from different eigenspaces are orthogonal.

\[
\det(\lambda I - A) = \begin{vmatrix}
\lambda - 2 & -1 - i \\
-1 + i & \lambda - 3
\end{vmatrix} = (\lambda - 2)(\lambda - 3) - (-1 - i)(-1 + i) = (\lambda - 1)(\lambda - 4)
\]

\[
\begin{bmatrix}
\lambda - 2 & -1 - i \\
-1 + i & \lambda - 3
\end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

\[
\begin{align*}
\lambda = 1: & \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} -1 - i \\ 1 \end{bmatrix} \\
\lambda = 4: & \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} \frac{1}{2}(1 + i) \\ 1 \end{bmatrix}
\end{align*}
\]

\[
v_1 = \begin{bmatrix} -1 - i \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{1}{2}(1 + i) \\ 1 \end{bmatrix} \quad \Rightarrow \quad v_1 \cdot v_2 = (-1 - i)\left(\frac{1}{2}(1 + i)\right) + (1)(1) = \frac{1}{2}(-1 - i)(1 - i) + 1 = 0
\]
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Theorem 8.9.6  If $A$ is an $n \times n$ matrix with complex entries, then the following are equivalent.

(a) $A$ is unitary.

(b) $\|Ax\| = \|x\|$ for all $x$ in $C^n$.

(c) $Ax \cdot Ay = x \cdot y$ for all $x$ and $y$ in $C^n$.

(d) The column vectors of $A$ form an orthonormal set in $C^n$ with respect to the complex Euclidean inner product.

(e) The row vectors of $A$ form an orthonormal set in $C^n$ with respect to the complex Euclidean inner product.
8.9. Hermitian, Unitary, and Normal Matrices

Hermitian and Unitary Matrices

Example 4

Use Theorem 8.9.6 show that

\[
A = \begin{pmatrix}
\frac{1}{2} (1 + i) & \frac{1}{2} (1 + i) \\
\frac{1}{2} (1 - i) & \frac{1}{2} (-1 + i)
\end{pmatrix}
\]

is unitary, and then find \( A^{-1} \).

\[
r_1 = \begin{pmatrix}
\frac{1}{2} (1 + i) \\
\frac{1}{2} (1 - i)
\end{pmatrix} \quad r_2 = \begin{pmatrix}
\frac{1}{2} (1 - i) \\
\frac{1}{2} (-1 + i)
\end{pmatrix}
\]

\[
\|r_1\| = \sqrt{\left| \frac{1}{2} (1 + i) \right|^2 + \left| \frac{1}{2} (1 + i) \right|^2} = \sqrt{\frac{1}{2} + \frac{1}{2}} = 1
\]

\[
\|r_2\| = \sqrt{\left| \frac{1}{2} (1 - i) \right|^2 + \left| \frac{1}{2} (-1 + i) \right|^2} = \sqrt{\frac{1}{2} + \frac{1}{2}} = 1
\]

\[
r_1 \cdot r_2 = \left( \frac{1}{2} (1 + i) \right) \left( \frac{1}{2} (1 - i) \right) + \left( \frac{1}{2} (1 + i) \right) \left( \frac{1}{2} (-1 + i) \right)
\]

\[
= \left( \frac{1}{2} (1 + i) \right) \left( \frac{1}{2} (1 + i) \right) + \left( \frac{1}{2} (1 + i) \right) \left( \frac{1}{2} (-1 - i) \right) = \frac{1}{2} i - \frac{1}{2} i = 0
\]

Since we now know that \( A \) is unitary, if follows that

\[
A^{-1} = A^* = \begin{pmatrix}
\frac{1}{2} (1 - i) & \frac{1}{2} (1 + i) \\
\frac{1}{2} (-1 - i) & \frac{1}{2} (-1 + i)
\end{pmatrix}
\]
8.9. Hermitian, Unitary, and Normal Matrices

Unitary Diagonalizability

Since unitary matrices are the complex analogs of the real orthogonal matrices, the following
definition is a natural generalization of the idea of orthogonal diagonalizability for real
matrices.

**Definition 8.9.7** A square complex matrix is said to be *unitarily diagonalizable* if there is
a unitary matrix $P$ such that $P^*AP = D$ is a complex diagonal matrix. Any such matrix $P$
is said to *unitarily diagonalize* $A$. 
8.9. Hermitian, Unitary, and Normal Matrices

Unitary Diagonalizability

**Theorem 8.9.8** Every $n \times n$ Hermitian matrix $A$ has an orthonormal set of $n$ eigenvectors and is unitarily diagonalized by any $n \times n$ matrix $P$ whose column vectors are an orthonormal set of eigenvectors of $A$.

Recall that a real symmetric $n \times n$ matrix $A$ has an orthonormal set of $n$ eigenvectors and is orthogonally diagonalized by any $n \times n$ matrix whose column vectors are an orthonormal set of eigenvectors of $A$.

1. Find a basis for each eigenspace of $A$.
2. Apply the Gram-Schmidt process to each of these bases to obtain orthonormal bases for the eigenspaces.
3. Form the matrix $P$ whose column vectors are the basis vectors obtained in the last step. This will be a unitary matrix (Theorem 8.9.6) and will unitarily diagonalize $A$. 
8.9. Hermitian, Unitary, and Normal Matrices

Unitary Diagonalizability

Example 5

Find a matrix $P$ that unitarily diagonalizes the Hermitian matrix

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

In Example 3, \( \lambda = 1: \)

$$v_1 = \begin{bmatrix} -1 - i \\ 1 \end{bmatrix} \quad \quad v_2 = \begin{bmatrix} \frac{1}{2} (1 + i) \end{bmatrix}$$

Since each eigenspace has only one basis vector, the Gram-Schmidt process is simply a matter of normalizing these basis vectors.

$$P = \begin{bmatrix} p_1 & p_2 \end{bmatrix} = \begin{bmatrix} \frac{-1 - i}{\sqrt{3}} & \frac{1 + i}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} \quad P^*AP = \begin{bmatrix} \frac{-1 + i}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1 - i}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 2 & 1 + i \\ 1 - i & 3 \end{bmatrix} \begin{bmatrix} \frac{-1 - i}{\sqrt{3}} & \frac{1 + i}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$
8.9. Hermitian, Unitary, and Normal Matrices

Skew-Hermitian Matrices

A square matrix with complex entries is **skew-Hermitian** if $A^* = -A$

A skew-Hermitian matrix must have zero or pure imaginary numbers on the main diagonal, and the complex conjugates of entries that are symmetrically positioned about the main diagonal must be negative of one another.

$$A = \begin{bmatrix} i & 1 - i & 5 \\ -1 - i & 2i & i \\ -5 & i & 0 \end{bmatrix}$$
8.9. Hermitian, Unitary, and Normal Matrices

Normal Matrices

Real symmetric matrices ↔ Orthogonally diagonalizable

Hermitian matrices ↔ Unitarily diagonalizable

It can be proved that a square complex matrix $A$ is unitarily diagonalizable if and only if $AA^* = A^*A$.

Matrices with this property are said to be normal.

Normal matrices include the Hermitian, skew-Hermitian, and unitary matrices in the complex case and the symmetric, skew-symmetric, and orthogonal matrices in the real case.
A Comparison of Eigenvalues
8.10. Systems of Differential Equations

Terminology

\[ y' = ay \implies y = ce^{at} \]

*general solution*

\[ y' = cae^{at} = a(ce^{at}) = ay \]

\[ y' = ay, \quad y(t_0) = y_0 \quad \text{initial value problem} \]

*initial condition*

**Example 1**

Solve the initial value problem

\[ y' = 2y, \quad y(0) = 6 \]

\[ y = ce^{2t} \implies y = 6e^{2t} \]
8.10. Systems of Differential Equations

Linear Systems of Differential Equations

\[ y'_1 = a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \]
\[ y'_2 = a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \]
\[ \vdots \quad \vdots \quad \vdots \]
\[ y'_n = a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n \]

\[
\begin{bmatrix}
  y'_1 \\
  y'_2 \\
  \vdots \\
  y'_n
\end{bmatrix}
= 
\begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix}
\]

\[ y' = Ay \quad (7) \]

*homogeneous first-order linear system*

\( A: \) coefficient matrix

Observe that \( y=0 \) is always a solution of \( (7) \). This is called the *zero solution* or the *trivial solution*. 
8.10. Systems of Differential Equations

Linear Systems of Differential Equations

Example 2

\[
\begin{bmatrix}
  y_1' \\
  y_2' \\
  y_3'
\end{bmatrix}
= \begin{bmatrix}
  3 & 0 & 0 \\
  0 & -2 & 0 \\
  0 & 0 & 5
\end{bmatrix}
\begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix}
\]

\[y_1(0) = 1, \quad y_2(0) = 4, \quad y_3(0) = -2\]

\[y_1' = 3y_1\]
\[y_2' = -2y_2 \quad \Rightarrow \quad y_1 = c_1 e^{3t}, \quad y_2 = c_2 e^{-2t}, \quad y_3 = c_3 e^{5t} \quad \Rightarrow \quad y = \begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} = \begin{bmatrix}
  c_1 e^{3t} \\
  c_2 e^{-2t} \\
  c_3 e^{5t}
\end{bmatrix}\]

\[y_3' = 5y_3\]

\[y(0) = \begin{bmatrix}
  y_1(0) \\
  y_2(0) \\
  y_3(0)
\end{bmatrix} = \begin{bmatrix}
  c_1 e^0 \\
  c_2 e^0 \\
  c_3 e^0
\end{bmatrix} = \begin{bmatrix}
  c_1 \\
  c_2 \\
  c_3
\end{bmatrix} = \begin{bmatrix}
  1 \\
  4 \\
  -2
\end{bmatrix} \quad \Rightarrow \quad y = \begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} = \begin{bmatrix}
  e^{3t} \\
  4e^{-2t} \\
  -2e^{5t}
\end{bmatrix}\]
8.10. Systems of Differential Equations

Fundamental Solutions

**Theorem 8.10.1**  *If* $y_1, y_2, \ldots, y_k$ *are solutions of* $y' = Ay$, *then

$$y = c_1y_1 + c_2y_2 + \cdots + c_ky_k$$

*is also a solution for every choice of the scalar constants* $c_1, c_2, \ldots, c_k$.

\[
y' = c_1y_1' + c_2y_2' + \cdots + c_ky_k' \\
y_1' = Ay_1, \quad y_2' = Ay_2, \ldots, \quad y_k' = Ay_k \\
\implies y' = c_1Ay_1 + c_2Ay_2 + \cdots + c_kAy_k = A(c_1y_1 + c_2y_2 + \cdots + c_ky_k) = Ay
\]

Every linear combination of solutions of $y' = Ay$ *is also a solution.*

Closed under addition and scalar multiplication $\implies$ *solution space*
8.10. Systems of Differential Equations

Fundamental Solutions

**Theorem 8.10.2**  If $A$ is an $n \times n$ matrix, then:

(a) The equation $y' = Ay$ has a set of $n$ linearly independent solutions.

(b) If $S = \{y_1, y_2, \ldots, y_n\}$ is any set of $n$ linearly independent solutions, then every solution can be expressed as a unique linear combination of the solutions in $S$.

We call $S$ a *fundamental set of solutions* of the system, and we call

$$y = c_1y_1 + c_2y_2 + \cdots + c_ny_n$$

a *general solution* of the system.
8.10. Systems of Differential Equations

Fundamental Solutions

Example 3

Find the fundamental set of solutions.

$$\begin{bmatrix} y_1' \\ y_2' \\ y_3' \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

In Example 2,

$$\mathbf{y} = \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-2t} \\ c_3 e^{5t} \end{bmatrix} = c_1 \begin{bmatrix} e^{3t} \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ e^{-2t} \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ e^{5t} \end{bmatrix} \rightarrow \text{A general solution}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} e^{3t} \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ e^{-2t} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ e^{5t} \end{bmatrix} \rightarrow \text{A fundamental set of solutions}$$
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

\[ y' = Ay \rightarrow y = e^{\lambda t} x \quad (17) \rightarrow y' = \lambda e^{\lambda t} x \]

vector \( x \) corresponds to coefficient \( c \)

\[ \lambda e^{\lambda t} x = Ae^{\lambda t} x \quad \Rightarrow \quad Ax = \lambda x \]

which shows that if there is a solution of the form (17), then \( \lambda \) must be an eigenvalue of \( A \) and \( x \) must be a corresponding eigenvector.

\[ \textbf{Theorem 8.10.3} \quad \text{If } \lambda \text{ is an eigenvalue of } A \text{ and } x \text{ is a corresponding eigenvector, then } y = e^{\lambda t} x \text{ is a solution of the system } y' = Ay. \]
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

Example 4

Use Theorem 8.10.3 to find a general solution of the system

\[ y'_1 = y_1 + y_2 \]
\[ y'_2 = 4y_1 - 2y_2 \]

\[ y' = Ay \quad A = \begin{bmatrix} 1 & 1 \\ 4 & -2 \end{bmatrix} \]

\[ \det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{vmatrix} = \lambda^2 + \lambda - 6 = (\lambda - 2)(\lambda + 3) \]

\[ \lambda = 2: \quad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \lambda = -3: \quad x_1 = \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix} \]

Therefore, \[ y_1 = e^{2t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad y_2 = e^{-3t} \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix} \]
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

Example 4

so a general solution of the system is

\[ y = c_1 e^{2t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix} \]

\[ y_1 = c_1 e^{2t} - \frac{1}{4} c_2 e^{-3t} \quad y_2 = c_1 e^{2t} + c_2 e^{-3t} \]
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

**Theorem 8.10.4**  If \( x_1, x_2, \ldots, x_k \) are linearly independent eigenvectors of \( A \), and if \( \lambda_1, \lambda_2, \ldots, \lambda_k \) are corresponding eigenvalues (not necessarily distinct), then

\[
y_1 = e^{\lambda_1 t} x_1, \quad y_2 = e^{\lambda_2 t} x_2, \ldots, \quad y_k = e^{\lambda_k t} x_k
\]

are linearly independent solutions of \( y' = Ay \).

If we assume that

\[
c_1 y_1 + c_2 y_2 + \cdots + c_k y_k = 0
\]

\[
\Rightarrow c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + \cdots + c_k e^{\lambda_k t} x_k = 0
\]

Setting \( t=0 \), we obtain \( c_1 x_1 + c_2 x_2 + \cdots + c_k x_k = 0 \)

As \( x_1, x_2, \ldots, x_k \) are linearly independent, so we must have

\[
c_1 = c_2 = \cdots = c_k = 0
\]

which proves that \( y_1, y_2, \ldots, y_k \) are linearly independent.
8.10. Systems of Differential Equations

**Solutions Using Eigenvalues and Eigenvectors**

**Theorem 8.10.5** If $A$ is a diagonalizable $n \times n$ matrix, then a general solution of the system $y' = Ay$ is

$$y = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + \cdots + c_n e^{\lambda_n t} x_n$$

(18)

where $x_1, x_2, \ldots, x_n$ are any $n$ linearly independent vectors and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues.

An $n \times n$ matrix is diagonalizable if and only if it has $n$ linearly independent eigenvectors (Theorem 8.2.6).
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

Example 5

Find a general solution of the system

\[ y_1' = -2y_3 \]
\[ y_2' = y_1 + 2y_2 + y_3 \]
\[ y_3' = y_1 + 3y_3 \]

\[ \Rightarrow y' = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} y \]

In Example 3 and 4 of Section 8.2, we showed the coefficient matrix is diagonalizable by showing that it has an eigenvalue \( \lambda = 1 \) with geometric multiplicity 1 and an eigenvalue \( \lambda = 2 \) with geometric multiplicity 2.

\[ p_1 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \text{ and } p_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad p_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \]

\[ \Rightarrow y = c_1 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} e^t + c_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} e^{2t} + c_3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} e^{2t} \]
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

Example 6

At time \( t=0 \), tank 1 contains 80 liters of water in which 7kg of salt has been dissolved, and tank 2 contains 80 liters of water in which 10kg of salt has been dissolved. Find the amount of salt in each tank at time \( t \).

Now let

\[ y_1(t) = \text{amount of salt in tank 1 at time } t \]
\[ y_2(t) = \text{amount of salt in tank 2 at time } t \]

\[ y_1' = \text{rate in-rate out} = \frac{y_2(t)}{8} - \frac{y_1(t)}{2} \]
\[ y_2' = \text{rate in-rate out} = \frac{y_1(t)}{2} - \frac{y_2(t)}{2} \]

\[ y_1' = -\frac{1}{2}y_1 + \frac{1}{8}y_2 \]
\[ y_2' = -\frac{1}{2}y_1 - \frac{1}{2}y_2 \]

\[ y_1(0) = 7, \quad y_2(0) = 10 \]
8.10. Systems of Differential Equations

Solutions Using Eigenvalues and Eigenvectors

Example 6

\[ A = \begin{bmatrix} -\frac{1}{2} & \frac{1}{8} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \]

\[ \lambda_1 = -\frac{3}{4}, \quad \mathbf{x}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \lambda_2 = -\frac{1}{4}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \]

\[ \mathbf{y} = c_1 e^{-3t/4} \mathbf{x}_1 + c_2 e^{-t/4} \mathbf{x}_2 \]

\[ y_1 = -c_1 e^{-3t/4} + c_2 e^{-t/4} \]
\[ y_2 = 2c_1 e^{-3t/4} + 2c_2 e^{-t/4} \]

\[ y_1(0) = 7, \quad y_2(0) = 10 \]

\[ -c_1 + c_2 = 7 \quad 2c_1 + 2c_2 = 10 \]

\[ y_1 = e^{-3t/4} + 6e^{-t/4} \]
\[ y_2 = -2e^{-3t/4} + 12e^{-t/4} \]

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
8.10. Systems of Differential Equations

Exponential Form of A Solution

\[ y' = Ay, \quad y(0) = y_0 \]

\[ y = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + \cdots + c_n e^{\lambda_n t} x_n = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \] (24)

where the constants \( c_1, c_2, \ldots, c_n \) are chosen to satisfy the initial condition.

\[ P = [x_1 \quad x_2 \quad \cdots \quad x_n] \] (25)

which is invertible and diagonalizes \( A \).

Setting \( t=0 \), we obtain

\[ y_0 = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = P \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \rightarrow \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = P^{-1} y_0 \] (26)
8.10. Systems of Differential Equations

Exponential Form of A Solution

\[ y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \rightarrow y = P \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} P^{-1} y_0 \]

\[ \rightarrow y = e^{tA} y_0 \]

Formula (21) of Section 8.3

**Theorem 8.10.6** If A is a diagonalizable matrix, then the solution of the initial value problem

\[ y' = Ay, \quad y(0) = y_0 \]

can be expressed as

\[ y = e^{tA} y_0 \] (27)
8.10. Systems of Differential Equations

Exponential Form of A Solution

Example 7

Use Theorem 8.10.6 to solve the initial value problem

\[ y_1' = -2y_3 \]
\[ y_2' = y_1 + 2y_2 + y_3 \quad y_1(0) = 2, \quad y_2(0) = -1, \quad y_3(0) = 0 \]
\[ y_3' = y_1 + 3y_3 \]

\[ A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}, \quad y(0) = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \]

\[ e^{tA} \] was computed in Example 5 of Section 8.3.

\[ y = e^{tA}y_0 = \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ e^{2t} - e^t & e^{2t} & e^{2t} - e^t \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4e^t - 2e^{2t} \\ e^{2t} - 2e^t \\ 2e^{2t} - 2e^t \end{bmatrix} \]
**8.10. Systems of Differential Equations**

The Case Where $A$ Is Not Diagonalizable

\[ y = e^{tA}y_0 \quad : \text{applicable in all cases} \]

Formula (21) of Section 8.3: applicable ONLY with diagonalizable matrices

If the coefficient matrix $A$ is not diagonalizable, the matrix $e^{tA}$ must be obtained or approximated using the infinite series

\[
e^{tA} = I + tA + \frac{t^2 A^2}{2!} + \frac{t^3 A^3}{3!} + \cdots + \frac{t^k A^k}{k!} + \cdots \quad (29)
\]
Exponential Form of A Solution

Example 8

Solve the initial value problem
\[
\begin{align*}
y_1' &= 0 \\
y_2' &= y_1 \\
y_3' &= y_1 + y_2
\end{align*}
\]
\[y_1(0) = 2, \quad y_2(0) = 1, \quad y_3(0) = 3\]

\[
A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad y(0) = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}
\]

\[A^3 = 0, \quad \text{from which it follows that (29) reduces to the finite sum}\]

\[
e^{tA} = I + tA + \frac{t^2A^2}{2} = \begin{bmatrix} 1 & 0 & 0 \\ t & 1 & 0 \\ t + \frac{1}{2}t^2 & t & 1 \end{bmatrix}
\]
8.10. Systems of Differential Equations

Exponential Form of A Solution

Example 8

\[
y = \begin{bmatrix}
1 & 0 & 0 \\
t & 1 & 0 \\
t + \frac{1}{2}t^2 & t & 1
\end{bmatrix}
\begin{bmatrix}
2 \\
1 \\
3
\end{bmatrix}
= \begin{bmatrix}
2 \\
1 + 2t \\
3 + 3t + t^2
\end{bmatrix}
\]

or equivalently,

\[
y_1 = 2, \quad y_2 = 1 + 2t, \quad y_3 = 3 + 3t + t^2
\]
8.10. Systems of Differential Equations

Exponential Form of A Solution

Example 9

Solve the initial value problem

\[
\begin{align*}
y'_1 &= y_2 \\
y'_2 &= -y_1
\end{align*}
\]

\[y_1(0) = 2, \quad y_2(0) = 1\]

\[
A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad y(0) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}
\]

\[A^2 = -I, \text{ from which it follows that} \]

\[
\begin{align*}
A^4 &= I, & A^6 &= -I, & A^8 &= I, & A^{10} &= -I, & \ldots \\
\end{align*}
\]

Thus, if we make the substitutions \(A^{2k} = (-1)^k I\) and \(A^{2k+1} = (-1)^k A\) (for \(k = 1, 2, \ldots\)) in (29) and use the Maclaurin series for \(\sin t\) and \(\cos t\), we obtain
8.10. Systems of Differential Equations

Exponential Form of A Solution

Example 9

\[ e^{tA} = I \left( 1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \frac{t^6}{6!} + \cdots \right) + A \left( t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} + \cdots \right) \]

\[ = I \cos t + A \sin t = \begin{bmatrix} \cos t & 0 \\ 0 & \cos t \end{bmatrix} + \begin{bmatrix} 0 & \sin t \\ -\sin t & 0 \end{bmatrix} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \]

\[ y = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \cos t + \sin t \\ \cos t - 2 \sin t \end{bmatrix} \]

or equivalently,

\[ y_1 = 2 \cos t + \sin t, \quad y_2 = \cos t - 2 \sin t \]