Chapter 5

EIGENVALUES, EIGENVECTORS, AND DIAGONALIZATION

5.1 THE IMPORTANCE OF DIAGONAL SIMILARITY; APPLICATION TO MARKOV CHAINS

Markov chains are a powerful tool for forecasting future events. Effective use of Markov chains involves the calculation of high powers of matrices. Such calculations can be tedious (by hand, at any rate). Extra information about the matrix can make the job easier, as Examples 1 and 2 show.

Example 1. Calculate $A^6$, where

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$$
Solution  By definition,

\[ A^6 = A \cdot A \cdot A \cdot A \cdot A \cdot A \]

\[
= \begin{pmatrix} -1 & 5 \\ -10 & 14 \end{pmatrix} \begin{pmatrix} -1 & 5 \\ -10 & 14 \end{pmatrix} \begin{pmatrix} -1 & 5 \\ -10 & 14 \end{pmatrix} \\
= \begin{pmatrix} -49 & 65 \\ -130 & 146 \end{pmatrix} \begin{pmatrix} -1 & 5 \\ -10 & 14 \end{pmatrix} \\
= \begin{pmatrix} -601 & 665 \\ -1330 & 1394 \end{pmatrix}
\]

**Example 2.** Calculate \( A^6 \), where

\[
A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}
\]

this time given that

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

where

\[
P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad P^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}
\]

Solution  Let

\[ D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \]

Since

\[ A^6 = PDP^{-1}PDP^{-1}PDP^{-1}PDP^{-1}PDP^{-1}PDP^{-1} = PD^6P^{-1} \]

and

\[
D^6 = \begin{pmatrix} 2^6 & 0 \\ 0 & 3^6 \end{pmatrix} = \begin{pmatrix} 65 & 0 \\ 0 & 729 \end{pmatrix}
\]

we see that

\[
A^6 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 65 & 0 \\ 0 & 729 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 64 & 729 \\ 64 & 1458 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \\
= \begin{pmatrix} -601 & 665 \\ -1330 & 1394 \end{pmatrix}
\]
5.1. **THE IMPORTANCE OF DIAGONAL SIMILARITY**

Example 2 shows that if a matrix $A$ is similar to a diagonal matrix $D$, then computing $A^n$ can be done by computing $D^n$, which is easy.

Two questions must be answered at this point:

1. Given $A$, can we find $P$ and $D$ so that $A = PDP^{-1}$?

2. How are powers of matrices used in Markov chains?

First, we answer question 2; question 1 is covered in Secs. 5.2 and 5.3. We begin with a simple discussion of Markov chains.

**Markov chains** are used to analyze **systems** which at a given time can be in only one of a finite number of states. For example, a person may or may not be in debt, the weather may be dry or wet, a mechanical system may be in equilibrium or not. In each case the system may change from state to state. In addition, there is a probability of transition from one state to another between successive observation times. The objective of Markov analysis is to calculate the probability that a system will be in a particular state at some future time and to determine the long-range behavior of the system.

**Definition 5.1.1.** A probability$^1$ is a number $p$ with $0 \leq p \leq 1$.

Intuitively, the **empirical probability** of an event $E$ in an experiment is

$$p = \frac{\text{number of occurrences of } E \text{ in a large number } N \text{ of experiments}}{N}$$

Notice that the minimum and maximum values of $p$ are 0 and 1, respectively.

**Example 3.** A coin is tossed 1000 times and allowed to land, and the result, heads or tails, is recorded. The observations are 700 heads, 300 tails. What is the empirical probability of the coin landing heads in a single toss? Is the coin a "fair" coin?

**Solution** The empirical probability of heads is $\frac{700}{1000} = .7$. The coin is not fair because by experience, the tossing of a fair coin would likely result in approximately 500 heads in 1000 flips.

The idea of the empirical probability makes possible a definition of the concept of a Markov chain.

---

$^1$We are using a brief intuitive notion of probability here; for a detailed development, see a text on probability theory.
**Definition 5.1.2.** Let a system $\mathcal{S}$ have possible states $s_1, s_2, \ldots, s_n$. Suppose that we observe $\mathcal{S}$ at given times $T_1, T_2, \ldots, T_m, \ldots$. A **Markov chain** is a process in which the empirical probability that $\mathcal{S}$ is in a particular state at observation time $T_k$ depends only on which state $\mathcal{S}$ is in at time $T_{k-1}$.

**Example 4.** Consider the system of a student $\mathcal{S}$ who is in states

- $s_1$: semester grade point average < 3.0
- $s_2$: semester grade point average ≥ 3.0

As a result of observations from high school, it is found that if $\mathcal{S}$ is in state $s_1$ in one semester, then she or he will work harder the next semester and achieve state $s_2$ with probability .8, state $s_1$ with the lower probability of .2. But if $\mathcal{S}$ is in state $s_2$ in one semester, then $\mathcal{S}$ relaxes the next semester and falls below 3.0 with probability .3; and $\mathcal{S}$ stays above 3.0 with probability .7. This is an example of a Markov chain with observation times at the end of each semester. We are assuming that $\mathcal{S}$’s achievement in one semester is determined only by her or his motivation resulting from the previous semester’s grade point average.

The probabilities in Example 4 are called **transition probabilities** and can be arranged in a matrix

$$
M = \begin{pmatrix}
.2 & .3 \\
.8 & .7
\end{pmatrix}
$$

Thus .8 is the transition probability of changing from state $s_1$ to $s_2$, and so on. Note that the columns add to 1. This matrix is called the **transition matrix** for the Markov chain.

The importance of the transition matrix is that it can be used to find the probability of being in state $s_1$ or $s_2$ at later times. To illustrate, suppose the student achieves a 2.5 grade point average (GPA) in the first semester. We know that the student is in state $s_1$ and can represent this with a state vector from $\mathcal{M}_{21}$ as

$$
S = \begin{pmatrix}
1 \\
0
\end{pmatrix}
$$
5.1. THE IMPORTANCE OF DIAGONAL SIMILARITY

The first slot is filled with the probability that \( S \) is in \( s_1 \) and the second slot with the probability she or he is in \( s_2 \). Multiplying \( S \) by \( M \), we have

\[
MS = \begin{pmatrix}
0.2 \\
0.8
\end{pmatrix}
\]

Recalling our example, we see that

\[
\begin{pmatrix}
0.2 \\
0.8
\end{pmatrix}
\]

represents the probabilities of being in \( s_1 \) and \( s_2 \) after a semester with GPA < 3.0. That is, \( MS \) is exactly the state vector for the next observation time. In general, this is true:

If \( S \), with transition matrix \( M \), has state vector \( S \) at time \( T_{k-1} \), then \( MS \) is the state vector for time \( T_k \).

Note that we require state vectors to have nonnegative entries which sum to 1.

**Example 5.** Show that if \( S \), with transition matrix \( M \), has state vector \( S \) at time \( T_0 \), then the state vector at time \( T_4 \) is \( M^4 S \).

**Solution** Using the statement above, the state vector at time \( T_1 \) is \( MS \). To find the state vector at time \( T_2 \), we multiply the state vector at time \( T_1 \) by \( M \). That is, we compute

\[
M( MS ) = M^2 S
\]

Continuing in this way, we have

<table>
<thead>
<tr>
<th>TIME</th>
<th>( T_0 )</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
<th>( T_3 )</th>
<th>( T_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>State vector</td>
<td>( S )</td>
<td>( MS )</td>
<td>( M( MS ) )</td>
<td>( M( M( MS ) ) )</td>
<td>( M( M( M( MS ) ) ) )</td>
</tr>
<tr>
<td>Simplified state vector</td>
<td>( S )</td>
<td>( MS )</td>
<td>( M^2 S )</td>
<td>( MS^3 )</td>
<td>( M^4 S )</td>
</tr>
</tbody>
</table>

From Example 5 we see the importance of the power of a matrix for Markov chains: It is used to compute future state vectors. The next example is more specific.
Example 6. Recall the student from Example 4. If the student achieves a 2.5 GPA in the first semester, what is the probability that \( S \) will have a GPA above 3.0 after the fourth semester?

Solution The initial state vector is

\[
\begin{pmatrix}
1 \\
0
\end{pmatrix}
\]

Schematically, we have the following:

<table>
<thead>
<tr>
<th>TIME</th>
<th>( t_0 )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semester</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
| State vector | \( \begin{pmatrix}
1 \\
0
\end{pmatrix} \) | \( M \begin{pmatrix}
1 \\
0
\end{pmatrix} \) | \( M^2 \begin{pmatrix}
1 \\
0
\end{pmatrix} \) | \( M^3 \begin{pmatrix}
1 \\
0
\end{pmatrix} \) |

So we see that

\[
M^3 \begin{pmatrix}
1 \\
0
\end{pmatrix}
\]

must be calculated. First, we calculate \( M^3 \),

\[
M = \begin{pmatrix}
.2 & .3 \\
.8 & .7
\end{pmatrix}
\]

\[
M^2 = \begin{pmatrix}
.28 & .27 \\
.72 & .73
\end{pmatrix}
\]

\[
M^3 = \begin{pmatrix}
.272 & .273 \\
.728 & .727
\end{pmatrix}
\]

Thus the state vector at the end of the fourth semester is

\[
M^3 S = \begin{pmatrix}
.272 \\
.728
\end{pmatrix}
\]

The probability of the student having a GPA above 3.0 in the fourth semester in .728. The student is much more likely to have a GPA above 3.0 than below.

Example 7. A utility company finds that, in general, if a typical customer pays a bill late one month, that person will pay before the due date on the next billing \( \frac{1}{3} \) of the time. But if a customer pays early one month, that person is likely to pay late the following month \( \frac{6}{10} \) of the time. For the March billing virtually all 10,000 customers pay their bills on time. About how many customers will pay late for the July billing?
5.1. THE IMPORTANCE OF DIAGONAL SIMILARITY

Solution We treat the payment system as a Markov chain. The states are

\[ s_1 = \text{payment on time} \quad s_2 = \text{payment late} \]

the transition matrix is

\[
M = \begin{pmatrix}
\frac{4}{10} & \frac{1}{2} \\
\frac{6}{10} & \frac{1}{2}
\end{pmatrix}
\]

This month

To find the state vector for July, we can make a chart

<table>
<thead>
<tr>
<th>MONTH</th>
<th>MARCH</th>
<th>APRIL</th>
<th>MAY</th>
<th>JUNE</th>
<th>JULY</th>
</tr>
</thead>
<tbody>
<tr>
<td>State vector</td>
<td>$S = \begin{pmatrix} 1 \ 0 \end{pmatrix}$</td>
<td>$MS$</td>
<td>$M^2S$</td>
<td>$M^3S$</td>
<td>$M^4S$</td>
</tr>
</tbody>
</table>

Or we could realize that July is month 7, March is month 3, and so we must calculate $M^{7-3} = M^4$. Calculating, we find

\[
M^4 = \begin{pmatrix}
.4546 & .4545 \\
.5454 & .5455
\end{pmatrix}
\]

Therefore the state vector is

\[
M^4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} .4546 \\ .5454 \end{pmatrix}
\]

The utility can expect about $.5454(10,000) = 5454$ people to pay late. The company probably needs to institute greater penalties for the late payment.

We discuss more properties of Markov chains in Secs. 5.2 and 5.3.

PROBLEMS 5.1

1. Given that

\[
A = \begin{pmatrix}
-1 & -3 \\
\frac{3}{2} & \frac{7}{2}
\end{pmatrix}
\begin{pmatrix}
2 & -1 \\
-1 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
0 & 2
\end{pmatrix}
\begin{pmatrix} 1 \\ 1 \end{pmatrix}
\]

calculate $A^{10}$. 
2. Which of the following matrices cannot be transition matrices for a Markov chain?

(a) \[
\begin{pmatrix}
-0.1 & 0.3 \\
0.9 & 0.7
\end{pmatrix}
\]  
(b) \[
\begin{pmatrix}
\frac{1}{2} & \frac{3}{2} \\
\frac{1}{2} & \frac{1}{2}
\end{pmatrix}
\]  
(c) \[
\begin{pmatrix}
\frac{1}{2} & \frac{1}{3} \\
\frac{1}{3} & \frac{2}{3}
\end{pmatrix}
\]  
(d) \[
\begin{pmatrix}
0.1 & 0.3 \\
0.7 & 0.1
\end{pmatrix}
\]  
(e) \[
\begin{pmatrix}
\frac{1}{2} & \frac{1}{3} & \frac{1}{10} \\
\frac{1}{3} & \frac{1}{3} & \frac{5}{10} \\
0 & \frac{1}{6} & \frac{4}{10}
\end{pmatrix}
\]  
(f) \[
\begin{pmatrix}
0.1 & 0.6 & 0.3 \\
0.7 & 0.4 & 0.1
\end{pmatrix}
\]

3. Consider a system of laundry detergent consumers. Consumers buy either liquid or dry detergent each week. During a previous advertising campaign for liquid detergents, a market research firm found that \(\frac{4}{10}\) of the people who bought dry detergent one week bought liquid the next. But \(\frac{8}{10}\) of the people who bought liquid one week bought dry the next week. Assume similar market dynamics in the new advertising campaign. If at the beginning of the campaign half the people buy liquid and half buy dry detergent, what share of the market will liquid detergent have after 4 weeks? Can this result be used to evaluate the advertising agency?

4. In a certain town it is known from past experience that if the weather is wet one day, there is a \(\frac{4}{10}\) probability of wet weather the next day. But if the weather is dry one day, then there is a \(\frac{9}{10}\) probability of dry weather the next day. On Monday the weather is wet, and a company’s picnic is scheduled for Saturday. Should the picnic be rescheduled?

5. Regarding Prob. 4, would an appreciable advantage be gained in having the picnic on Sunday instead of Saturday?

### 5.2 THE EIGENPROBLEM

We saw in Sec. 5.1 that it is important to know when we can write a matrix \(A\) as

\[
A = PDP^{-1}
\]
where $D$ is a diagonal matrix. To see where $D$ comes from, let us suppose that $A$ is $2 \times 2$,

$$A = P \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} P^{-1}$$

and that $P$ represents a rotation of $\theta$ radians in $E^2$. Let $X$ in $M_{21}$ be a vector that $P^{-1}$ rotates to

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{or} \quad P^{-1}X = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Then

$$AX = P \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = P \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 2P \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2X$$

(We know that

$$P \begin{pmatrix} 1 \\ 0 \end{pmatrix} = X$$

since $P$ just undoes the rotation $P^{-1}$.) We see then that $AX = 2X$. In this case we say the number 2 is an eigenvalue (pronounced “eye-genvalue”) of $A$ and $X$ is an eigenvector of $A$. We can also see that 3 is an eigenvalue of $A$ by letting $Y$ be a vector which $P$ rotates to

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then

$$AY = P \begin{pmatrix} 0 \\ 3 \end{pmatrix} = 3P \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 3Y$$

The calculations in the last paragraph showed that for the given matrix, the entries of $D$ satisfied the equations $AX = 2X$ and $AY = 3Y$. This suggests that the following problem, the fourth basic problem of the linear algebra, is important in finding diagonal matrices similar to a given matrix.

**The Matrix Eigenproblem**

Given an $n \times n$ matrix $A$ in $C_{nm}$, find all numbers $\lambda$ and all nonzero vectors $X$ in $C_{n1}$ such that $AX = \lambda X$. The numbers
\( \lambda \) are called eigenvalues\(^2\) of \( A \), and the vectors \( X \) are called eigenvectors\(^2\) of \( A \).

The solution of the matrix eigenproblem reduces to the solution of linear equations (the first basic problem of linear algebra), as Example 1 shows.

**Example 1.** Find the eigenvalues and eigenvectors of

\[
A = \begin{pmatrix}
1 & 1 \\
-2 & 4
\end{pmatrix}
\]

**Solution** We must solve \( AX = \lambda X \). In expanded form this is

\[
\begin{pmatrix}
1 & 1 \\
-2 & 4
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = \lambda \begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = \begin{pmatrix}
\lambda x_1 \\
\lambda x_2
\end{pmatrix}
\]

Reducing this to linear equations, we find

\[
(A - \lambda I)X = \begin{pmatrix}
1 - \lambda & 1 \\
-2 & 4 - \lambda
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} - \begin{pmatrix}
\lambda & 0 \\
0 & \lambda
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = \begin{pmatrix}
0 \\
0
\end{pmatrix}
\]

or

\[
(1 - \lambda)x_1 + x_2 = 0 \\
-2x_1 + (4 - \lambda)x_2 = 0
\]

(5.2.1)

Since we have only two equations and three unknowns \( (\lambda, x_1, x_2) \), we expect a third condition to make solution possible. Because eigenvectors are required to be nonzero, we realize that the system (5.2.1) must have a nontrivial solution for

\[
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
\]

From our previous work on homogeneous equations, we know that a nontrivial solution exists if and only if

\[
\det \begin{pmatrix}
1 - \lambda & 1 \\
-2 & 4 - \lambda
\end{pmatrix} = 0
\]

This condition reduces to

\[
(1 - \lambda)(4 - \lambda) + 2 = 6 - 5\lambda + \lambda^2 = (\lambda - 3)(\lambda - 2) = 0
\]

\(^2\)They are also called characteristic values and characteristic vectors. In many applications, the matrices, eigenvalues, and eigenvectors are required to be real.
which means that $\lambda = 2$ and $\lambda = 3$ are solutions to $AX = \lambda X$, with $X$ still to be found.

To find $X$ corresponding to $\lambda = 2$, substitute $\lambda = 2$ into (5.2.1), to obtain

$$-x_1 + x_2 = 0$$
$$-2x_1 + 2x_2 = 0$$

which has solution

$$\begin{pmatrix} k \\ k \end{pmatrix}$$

Thus the general eigenvector corresponding to the eigenvalue 2 is

$$\begin{pmatrix} k \\ k \end{pmatrix}$$

Specific eigenvectors are obtained by putting $k$ equal to specific numbers.

To calculate $X$ corresponding to $\lambda = 3$, substitute $\lambda = 3$ into (5.2.1) to find

$$-2x_1 + x_2 = 0$$
$$-2x_1 + x_2 = 0$$

This has solution

$$\begin{pmatrix} j \\ 2j \end{pmatrix}$$

which is the general eigenvector of $A$ corresponding to the eigenvalue 3.

Finally, we write the solution to the eigenproblem for $A$ by listing the eigenpairs

$$\begin{pmatrix} 2, \begin{pmatrix} k \\ k \end{pmatrix} \end{pmatrix}, \quad \begin{pmatrix} 3, \begin{pmatrix} j \\ 2j \end{pmatrix} \end{pmatrix}$$

Solution procedure for the eigenproblem for $A_{n \times n}$

1. Solve $\det(A - \lambda I) = 0$ for eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

2. To find the general eigenvector corresponding to $\lambda$, solve the homogeneous equations $(A - \lambda I)X = 0$.

Closely related to the matrix eigenproblem is the eigenproblem for linear transformations.
Given $L : V \rightarrow V$, a linear transformation, find all numbers $\lambda$ and nonzero vectors $v$ in $V$ such that

$$L(v) = \lambda v$$

To solve this problem, we need to solve the matrix eigenproblem for a matrix which represents $L$. We can solve the problem this way because all matrices representing $L$ have the same eigenvalues. This follows from Theorem 5.2.1 and 5.2.2.

**Theorem 5.2.1.** If $A$ is an $n \times n$ matrix, then

(a) $\text{Det} (A - \lambda I)$ is a polynomial $p(\lambda)$ of degree $n$.

(b) The eigenvalues of $A$ are the solutions of $p(\lambda) = 0$.

(c) If $\lambda_0$ is an eigenvalue, any nontrivial solution of $(A - \lambda_0 I)X = 0$ is an eigenvector of $A$ corresponding to $\lambda_0$.

**Proof.** (a) If $A$ is $2 \times 2$, then $\text{det}(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21}$, which is a polynomial in $\lambda$ of degree 2. To proceed by induction, suppose that for a $k \times k$ matrix the determinant of $A - \lambda I$ is a polynomial of degree $k$. Now if $B$ is a $(k + 1) \times (k + 1)$ matrix, then we calculate $\text{det}(B - \lambda I)$ by using the definition and column 1. In this way the determinant is seen to be a polynomial of degree $k + 1$, and the principle of mathematical induction proves part (a).

(b) A number $\lambda$ is an eigenvalue if and only if $AX = \lambda X$ and $X \neq 0$. Thus, $AX - \lambda X = (A - \lambda I)X = 0$. Homogeneous equations have nontrivial solutions if and only if the determinant of the coefficient matrix is zero. Therefore $\lambda$ is an eigenvalue if and only if $p(\lambda) = \text{det}(A - \lambda I) = 0$.

(c) Let $X$ be any nontrivial solution of $(A - \lambda_0 I)X = 0$. Then $AX = \lambda_0 IX = \lambda_0 x$.

The polynomial $p(\lambda)$ mentioned in Theorem 5.2.1 is called the characteristic polynomial of $A$. If $A$ is $n \times n$, its characteristic polynomial has $n$ roots. Therefore an $n \times n$ matrix has $N$ eigenvalues.

**Theorem 5.2.2.** If $A$ and $B$ are similar, then the set of eigenvalues of $A$ is equal to the set of eigenvalues of $B$. If the similarity transform is $P$ (that is, if $A = P^{-1}BP$) and $(\lambda_0, X)$ is an eigenpair of $A$, then $(\lambda_0, PX)$ is an eigenpair of $B$. 
5.2. THE EIGENPROBLEM

Proof. Let $A$ be similar to $B$, and let $\lambda_0$ be an eigenvalue of $A$. We have, using $A = P^{-1}BP$,

$$0 = \det(A - \lambda_0 I) = \det(P^{-1}BP - \lambda_0 I)$$
$$= \det(P^{-1}BP - \lambda_0 P^{-1}IP)$$
$$= \det(P^{-1}BP - P^{-1}(\lambda_0 I)P)$$
$$= \det(P^{-1}(B - \lambda_0 I)P)$$
$$= \det(P^{-1}) \det(B - \lambda_0 I) \det P$$
$$= \det(B - \lambda_0 I)$$

Thus, $\lambda_0$ is an eigenvalue of $B$. So every eigenvalue of $A$ is also an eigenvalue of $B$. The argument is virtually the same to show that every eigenvalue of $B$ is an eigenvalue of $A$.

Now let $(\lambda_0, X)$ be an eigenpair of $A$. We have

$$AX = \lambda_0 X$$

so

$$(P^{-1}BP)X = \lambda_0 X$$

Multiplying both sides on the left by $P$ and reassociating the products lead to

$$B(PX) = \lambda_0 (PX)$$

Now we can state a solution to the eigenproblem for linear transformations.

Choose a basis $\mathcal{S}$ to $V$. Let $M$ be the matrix representing $L$ with respect to the basis $\mathcal{S}$. If $(\lambda_0, X)$ is an eigenpair of $M$, then letting $\mathbf{v}$ be the vector such that $(\mathbf{v})_\mathcal{S} = X$, we have

$$L(\mathbf{v}) = \lambda_0 \mathbf{v}$$

Because of this correspondence we work mainly with the matrix eigenvalue problem. Now we consider some more examples of this problem.
Example 2. Show that

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

has complex eigenvalues.

Solution  We have

\[ \det(A - \lambda I) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1 \]

The equation \( \det(A - \lambda I) = 0 \) is \( \lambda^2 + 1 = 0 \), which has solutions \( \lambda_1 = i \) and \( \lambda_2 = -i \).

For the eigenvectors we solve \( (A - \lambda I)X = 0 \) for \( X \) when \( \lambda \) is \( i \) or \( -i \).

The equations are

\[ \lambda_1 = i: \quad \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \xrightarrow{R_1 + R_2} \begin{pmatrix} -i & 1 \\ 0 & 0 \end{pmatrix} \]
\[ x_2 = k \quad x_1 = \frac{1}{i} k = -ik \]

\[ X = k \begin{pmatrix} -i \\ 1 \end{pmatrix} \]

\[ \lambda_2 = -i: \]

\[
\begin{pmatrix}
  i & 1 & 0 \\
-1 & i & 0
\end{pmatrix}
\xrightarrow{-iR_1 + R_2}
\begin{pmatrix}
i & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

\[ x_2 = r \quad x_1 = \frac{-1}{i} r = ir \]

\[ X = r \begin{pmatrix} i \\ 1 \end{pmatrix} \]

Therefore the solution to the eigenproblem is

\[
\begin{pmatrix}
i, c \begin{pmatrix}-i \\ 1\end{pmatrix} \\
-i, d \begin{pmatrix}i \\ 1\end{pmatrix}
\end{pmatrix}
\]

where \( c \) and \( d \) are arbitrary constants.

In the first two examples the eigenvalues were distinct. However, this need not always be the case.

**Example 3.** Solve the eigenproblem for

\[
A = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}
\]

**Solution** Solving \( \det(A - \lambda I) = (3 - \lambda)^2 = 0 \), we find \( \lambda = 3 \) is a solution of multiplicity 2 since \( 3 - \lambda \) is twice a factor of \( \det(A - \lambda I) \). To find the eigenvector corresponding to \( \lambda = 3 \), we substitute \( \lambda = 3 \) into \( (A - \lambda I)X = 0 \) to obtain

\[
0x_1 + 0x_2 = 0 \\
0x_1 + 0x_2 = 0
\]

which means that any vector \( X \) is an eigenvalue corresponding to \( \lambda = 3 \).

That is, the eigenpairs are

\[
\begin{pmatrix} 3, \begin{pmatrix} k \\ j \end{pmatrix} \end{pmatrix} \quad \begin{pmatrix} 3, \begin{pmatrix} r \\ s \end{pmatrix} \end{pmatrix}
\]
where \( r, s, k, \) and \( j \) are arbitrary.

Before proceeding to more examples, we summarize what we have done so far.

An eigenvector of a matrix \( A \) is a number \( \lambda \) which satisfies the equation \( \det(A - \lambda I) = 0 \). An eigenvector corresponding to \( \lambda \) is any nonzero solution of \((A - \lambda I)X = 0\). The equation \( \det(A - \lambda I) = 0 \) is called the characteristic equation of \( A \). The expression \( \det(A - \lambda I) \) is always a polynomial (of degree \( n \) if \( A \) is \( n \times n \)) and is called the characteristic polynomial of \( A \). The eigenvalues of \( A \) are the roots of the characteristic polynomial and the solutions of the characteristic equation. From algebra we know that a polynomial of degree \( n \) has \( n \) complex roots; therefore an \( n \times n \) matrix \( A \) has \( n \) eigenvalues. In counting eigenvalues, multiplicities must be taken into account. For instance, in Example 3, the \( 2 \times 2 \) matrix had two eigenvalues, because the root, 3, of the characteristic polynomial was a repeated root.

If \((\lambda_0, X)\) and \((\lambda_0, Y)\) are two eigenpairs of \( A \), then \((\lambda_0, X + Y)\) and \((\lambda_0, cX)\) are eigenpairs of \( A \), provided \( X + Y \neq 0 \) and \( c \neq 0 \). Therefore, if we adjoin the zero vector to the set of all eigenvectors associated with \( \lambda_0 \), we obtain a subspace of \( \mathcal{C}_{n1} \). This subspace is denoted \( E_{(\lambda_0)} \), and is called the eigenspace of \( \lambda_0 \). Thus for matrix \( A \) in Example 1,

\[
E_{(2)} = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad \text{and} \quad E_{(3)} = \text{span} \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}
\]

**Example 4.** Show that

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

has only one real eigenvalue. Find the eigenspace corresponding to that eigenvalue.

**Solution** The characteristic polynomial is

\[
\det \begin{pmatrix}
-\lambda & 1 & 0 \\
-1 & -\lambda & 0 \\
0 & 0 & 2 - \lambda
\end{pmatrix} = (\lambda^2 + 1)(2 - \lambda)
\]

Since \( \lambda^2 + 1 \) is never zero for real \( \lambda \), \( \lambda = 2 \) is the only real eigenvalue. The
eigenvector is found by solving $A - \lambda I = 0$ with $\lambda = 2$:

$$\begin{pmatrix} -2 & 1 & 0 & 0 \\ -1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The general solution is

$$\begin{pmatrix} 0 \\ 0 \\ k \end{pmatrix}$$

and the eigenpair is

$$\begin{pmatrix} 2, \begin{pmatrix} 0 \\ 0 \\ k \end{pmatrix} \end{pmatrix}$$

The eigenspace

$$E_{(2)} = \text{span} \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Example 5. Solve the eigenproblem for

$$A = \begin{pmatrix} -4 & -4 & -8 \\ 4 & 6 & 4 \\ 6 & 4 & 10 \end{pmatrix}$$

Solution  The characteristic polynomial $p(\lambda)$ is

$$\det \begin{pmatrix} -4 - \lambda & -4 & -8 \\ 4 & 6 - \lambda & 4 \\ 6 & 4 & 10 - \lambda \end{pmatrix} = -\lambda^3 + 12\lambda^2 - 44\lambda + 48$$

To solve $p(\lambda) = 0$, we first guess at one root of $p(\lambda)$. First we try integer factors of 48: $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 8, \pm 12, \pm 16, \pm 24, \pm 48$. Since $p(1) \neq 0$ and $p(-1) \neq 0$, neither 1 nor $-1$ is a root. However, $p(2) = 0$, so 2 is a root
and $\lambda - 2$ is a factor. Dividing, we have

\[
-\lambda^2 + 10\lambda - 24 \\
\lambda - 2\sqrt{-\lambda^2 + 12\lambda^2 - 44\lambda + 48} \\
\frac{-\lambda^3 + 2\lambda^2}{10\lambda^2 - 44\lambda} \\
\frac{-\lambda^3 + 2\lambda^2}{10\lambda^2 - 20\lambda} \\
-24\lambda + 48 \\
-24\lambda + 48 \\
0
\]

so that

\[
p(\lambda) = (\lambda - 2)(-\lambda^2 + 10\lambda - 24)
\]

Now we can factor the quadratic part of $p(\lambda)$ easily into $(-\lambda + 4)(\lambda - 6)$. Therefore

\[
p(\lambda) = (\lambda - 2)(-\lambda + 4)(\lambda - 6)
\]

and the eigenvalues are 2, 4, and 6.

To find the eigenvectors, we substitute $\lambda = 2, 4, 6$ into $(A - \lambda I)X = 0$ and solve. The resulting equations are

\[
\begin{array}{l}
\lambda = 2: \quad \begin{pmatrix} -6 & -4 & -8 & 0 \\ 4 & 4 & 4 & 0 \\ 6 & 4 & 8 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 3 & 2 & 4 & 0 \\ 0 & 4 & -4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow X = \begin{pmatrix} -2k \\ k \end{pmatrix} \\
\lambda = 4: \quad \begin{pmatrix} -8 & -4 & -8 & 0 \\ 4 & 2 & 4 & 0 \\ 6 & 4 & 6 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 & 2 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow X = \begin{pmatrix} -j \\ 0 \end{pmatrix} \\
\lambda = 6: \quad \begin{pmatrix} -10 & -4 & -8 & 0 \\ 4 & 0 & 4 & 0 \\ 6 & 4 & 4 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} -5 & -2 & -4 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow X = \begin{pmatrix} -r \\ \frac{1}{2}r \end{pmatrix}
\end{array}
\]

The eigenpairs are

\[
\begin{pmatrix} 2, \begin{pmatrix} -2k \\ k \end{pmatrix} \end{pmatrix}, \begin{pmatrix} 4, \begin{pmatrix} -j \\ 0 \end{pmatrix} \end{pmatrix}, \begin{pmatrix} 6, \begin{pmatrix} -r \\ \frac{1}{2}r \end{pmatrix} \end{pmatrix}
\]
and the eigenspaces are

\[ E_{(2)} = \text{span} \left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \right\} \quad E_{(4)} = \text{span} \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} \quad E_{(6)} = \text{span} \left\{ \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix} \right\} \]

**Example 6.** The matrices

\[
A = \begin{pmatrix}
3 & -1 & 0 \\
0 & 3 & 0 \\
0 & 0 & 5
\end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix}
3 & 0 & 0 \\
0 & 3 & 0 \\
0 & 0 & -5
\end{pmatrix}
\]

both have eigenvalues 3 and -5, with 3 being an eigenvalue of multiplicity 2. Compare the eigenspaces for these two matrices.

**Solution** For A the general eigenpairs are

\[
\left( 3, \begin{pmatrix} k \\ 0 \\ 0 \end{pmatrix} \right) \quad \text{and} \quad \left( -5, \begin{pmatrix} 0 \\ 0 \\ r \end{pmatrix} \right)
\]

so the eigenspaces are

\[ E_{(3)} = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\} \quad E_{(-5)} = \text{span} \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \]

For B the general eigenpairs are

\[
\left( 3, \begin{pmatrix} r \\ s \\ 0 \end{pmatrix} \right) \quad \text{and} \quad \left( -5, \begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix} \right)
\]

so the eigenspaces for B are

\[ E_{(3)} = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \quad E_{(-5)} = \text{span} \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \]

For A, \( \dim E_{(3)} < \) (multiplicity of \( \lambda = 3 \)), while for B, \( \dim E_{(3)} = \) (multiplicity of \( \lambda = 3 \)).
In the preceding examples we have been able to solve the eigenproblem exactly. In general, this is not the case; sometimes numerical methods must be used. These are discussed in Chap. 6.

**Fixed Vectors of Matrices** If a matrix $A$ has an eigenpair $(1, X)$, (that is, if $AX = X$), then $X$ is called a **fixed vector** (or fixed point) of $A$. This concept is important for Markov chains. Some Markov chains have the property that as $n$ grows larger and larger, $M^n$ (powers of the transition matrix) begins to look the same. For example, the matrix

$$M = \begin{pmatrix} .2 & .3 \\ .8 & .7 \end{pmatrix}$$

from Example 5 of Sec. 5.1 has powers

$$M^2 = \begin{pmatrix} .28 & .27 \\ .72 & .73 \end{pmatrix}$$
$$M^3 = \begin{pmatrix} .272 & .273 \\ .728 & .727 \end{pmatrix}$$
$$M^4 = \begin{pmatrix} .2728 & .2727 \\ .7272 & .7273 \end{pmatrix}$$
$$M^5 = \begin{pmatrix} .27272 & .27273 \\ .72728 & .72727 \end{pmatrix}$$
$$M^6 = \begin{pmatrix} .272728 & .272727 \\ .727272 & .727273 \end{pmatrix}$$

which appear to be approximating a matrix

$$T = \begin{pmatrix} .2727 \cdots & .2727 \cdots \\ .7272 \cdots & .7272 \cdots \end{pmatrix}$$

$$= \begin{pmatrix} \frac{27}{99} & \frac{27}{99} \\ \frac{72}{99} & \frac{72}{99} \end{pmatrix}$$

(see Prob. 5). In this case we write

$$M^n \to T \quad \text{as} \quad n \to +\infty$$

If a Markov chain has this property, it is called **regular**. In a regular Markov chain, for any initial state vector $S$

$$M^n S \to TS$$
5.2. THE EIGENPROBLEM

This means that regardless of initial state, the Markov chain settles into an
**equilibrium state** \( E = TS \).

The important fact is this:

If a Markov chain with transition matrix \( M \) is regular, then the
equilibrium state \( E \) is a fixed point of \( M \). That is,

\[
ME = E
\]

**Example 7.** It can be shown that a Markov chain is **regular** if its transition
matrix \( M \), or some power of \( M \), has only positive entries. Consider a Markov
chain with transition matrix

\[
M = \begin{pmatrix}
.3 & 1 \\
.7 & 0
\end{pmatrix}
\]

Show that the chain is regular. Find the equilibrium state vector for the
chain.

**Solution** Since \( M \) does not have all positive entries, we must look at powers
of \( M \) to establish regularity. Since

\[
M^2 = \begin{pmatrix}
.77 & 3 \\
.21 & 7
\end{pmatrix}
\]

has all positive entries, the chain is regular. To find the fixed point \( E \), we
must solve

\[
ME = E
\]

with \( E \) being a state vector. That is, we must solve

\[
ME = E
\]

\[
E^+ = 1 \quad E \geq 0
\]

where \( E \geq 0 \) means all the components of \( E \) are nonnegative and \( E^+ \) means
the sum of the components of \( E \).

First, we solve \( ME = E \). Letting

\[
ME = E = \begin{pmatrix}
e_1 \\
e_2
\end{pmatrix}
\]
we have
\[
\begin{pmatrix}
.3 & 1 \\
.7 & 0
\end{pmatrix}
\begin{pmatrix}
e_1 \\
e_2
\end{pmatrix} =
\begin{pmatrix}
e_1 \\
e_2
\end{pmatrix}
\]
which reduces to
\[-.7e_1 + e_2 = 0 \quad .7e_1 - e_2 = 0\]
or \(-.7e_1 + e_2 = 0\). Therefore
\[
E = \begin{pmatrix}
\frac{10}{17} k \\
k
\end{pmatrix}
\]
The requirement \(E^+ = 1\) forces the condition
\[
\frac{10}{17} k + k = 1 \\
k = \frac{17}{17}
\]
do that
\[
E = \begin{pmatrix}
\frac{10}{17} \\
\frac{17}{17}
\end{pmatrix}
\]
is the equilibrium state for the chain. The interpretation of this result is that if \(M\) is the transition matrix for a Markov chain, then in the long run the system is in state number 1 for \(\frac{10}{17}\) of the time and in state number 2 for \(\frac{7}{17}\) of the time. Note that the condition \(E^+ = 1\) forced us to choose only one of the eigenvectors with eigenvalue 1.

**Example 8.** Show that the matrix
\[
M = \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
cannot be the transition matrix of a regular Markov chain.

**Solution** We will show that powers of \(M\) do not tend to a particular matrix. The powers are
\[
M^2 = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \quad M^3 = \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix} \quad M^4 = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \quad \ldots
\]
Thus
\[
M^{2n} = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \quad M^{2n+1} = \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
and the associated Markov chain is not regular.
5.2. THE EIGENPROBLEM

Example 9. A robotics company is to manufacture a robot arm which attempts to pick parts off one conveyer belt and put them on another. The arm occasionally fails to grasp the part securely and does not transfer the part to the second belt successfully. The robot is designed so that in case of failure it “tries harder” in the sense that secondary circuits are activated. In trials it is found that if the arms fails at one time, it succeeds the next time 97 percent of the time. If the arms succeeds at a given time, the secondary circuits are deactivated and the arm will fail the next time only 2 percent of the time. Will the arm satisfy the customer’s requirement that it work successfully 98 percent of the time?

Solution The transition matrix is

\[
M = \begin{pmatrix}
.98 & .97 \\
.02 & .03
\end{pmatrix}
\]

To find the fixed point \( ME = E, \ E^+ = 1, \ E \geq 0, \) we solve

\[
-0.02x_1 + 0.97x_2 = 0 \\
0.02x_1 - 0.97x_2 = 0
\]

to find

\[
X = \begin{pmatrix}
48.5k \\
k
\end{pmatrix}
\]

The condition \( E^+ = 1 \) forces \( k = 0.02 \overline{02} \), so that

\[
E = \begin{pmatrix}
0.97\overline{97} \\
0.02\overline{02}
\end{pmatrix}
\]

In the long run the arm will work successfully 97.97 percent of the time. Strictly speaking, the arm is not satisfactory; however, 97.97 percent is very close to 98 percent, and perhaps the customer would accept it.

Example 10. A space vehicle has three navigational computers. Each onboard navigation computer constantly performs internal checks on its own
circuits. If a circuit has failed, the computer will cease navigation functions and repair itself. The computer is therefore in one of three states

\[ s_1 \text{ performing navigation functions} \]
\[ s_2 \text{ failed, repair not begun} \]
\[ s_3 \text{ failed, repairing itself} \]

In trials the transition matrix is

\[
M = \begin{pmatrix}
1 & 0 & 0 \\
0.98 & 0 & 0.9 \\
0.01 & 1 & 0 \\
0.01 & 0.9 & 0.1
\end{pmatrix}
\]

Approximately what percentage of the time is the computer operational?

**Solution** First, by checking \( M^2 \) we find that the chain is regular. Solving \( ME = E, E^+ = 1, E \geq 0 \), we obtain

\[
E = \begin{pmatrix}
0.96774 \cdots \\
0.01075 \cdots \\
0.021505 \cdots
\end{pmatrix}
\]

The conclusion is that the computer is operational about 96.77 percent of the time. By using laws of probability which we have not discussed, the probability that all three computers would be in failure at the same time is less than 0.000034 percent.

Although Markov chains are not the only application in which fixed vectors of matrices are used, they are one of the most important.

**PROBLEMS 5.2**

1. Solve the eigenproblem for the following matrices.
5.2. THE EIGENPROBLEM

(a) \[
\begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix}
\]
(b) \[
\begin{pmatrix}
1 & 0 \\
1 & 1
\end{pmatrix}
\]
(c) \[
\begin{pmatrix}
0 & -1 \\
-1 & 0
\end{pmatrix}
\]
(d) \[
\begin{pmatrix}
1 & 1 & 0 \\
0 & -1 & 1 \\
0 & 0 & 2
\end{pmatrix}
\]
(e) \[
\begin{pmatrix}
2 & -1 \\
-4 & 2
\end{pmatrix}
\]
(f) \[
\begin{pmatrix}
1 & -1 & 4 \\
3 & 2 & -1 \\
2 & 1 & -1
\end{pmatrix}
\]
(g) \[
\begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & -2 & 4
\end{pmatrix}
\]
(h) \[
\begin{pmatrix}
0 & i \\
i & 0
\end{pmatrix}
\]
(i) \[
\begin{pmatrix}
i & 1 \\
0 & i
\end{pmatrix}
\]
(j) \[
\begin{pmatrix}
0 & -i \\
i & 0
\end{pmatrix}
\]

2. Find fixed points (if they exist) of the following matrices.

(a) \[
\begin{pmatrix}
\frac{1}{2} & \frac{3}{2} \\
\frac{3}{2} & \frac{1}{2}
\end{pmatrix}
\]
(b) \[
\begin{pmatrix}
0 & -1 \\
-1 & 0
\end{pmatrix}
\]
(c) \[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
(d) \[
\begin{pmatrix}
2 & 3 \\
0 & -1
\end{pmatrix}
\]
(e) \[
\begin{pmatrix}
-1 & 0 & 1 \\
0 & 2 & -1 \\
0 & 0 & 3
\end{pmatrix}
\]
(f) \[
\begin{pmatrix}
2 & 0 & 0 \\
1 & 1 & 0 \\
7 & 3 & 0
\end{pmatrix}
\]

3. For Prob. 3 of Sec. 5.1 find the equilibrium state for the Markov chain.

4. For Prob. 4 of Sec. 5.1 find the equilibrium state for the Markov chain.

5. Show that \(0.2727\cdots = \frac{27}{99}\) by writing

\[
0.2727\cdots = 0.27 + 0.0027 + \cdots = \frac{27}{99} + \frac{27}{9900} + \cdots
\]

\[
= \frac{27}{99} + \frac{27}{9900} \left(\frac{1}{100}\right) + \frac{27}{9900} \left(\frac{1}{100}\right)^2 + \cdots
\]

and using the formula for the sum of

\[a + ar + ar^2 + \cdots \quad |r| < 1\]

which is \(a/(1 - r)\). Do the same for \(0.7272\cdots = \frac{72}{99}\).
6. Show that if $X$ is an eigenvector of $A$, then so is $kX$ for any $k \neq 0$. [**Hint:** Look at $A(kX)$.] 

7. Show that if $\lambda$ is an eigenvalue of $A$, then $k\lambda$ is an eigenvalue of $kA(k \neq 0)$. 

8. Show that if $\lambda$ is an eigenvalue of $A$, then $\lambda^2$ is an eigenvalue of $A^2$. [**Hint:** Look at $A(AX)$.] 

9. Show that if $\lambda \neq 0$ is an eigenvalue of $A$ and if $A^{-1}$ exists, then $1/\lambda$ is an eigenvalue of $A^{-1}$. 

10. Show that if $\lambda = 0$ is an eigenvalue of $A$, then $A$ is singular. [**Hint:** $\lambda$ must satisfy $\det(A - \lambda I) = 0$.] 

11. Show that the eigenvalues of a triangular matrix are just the diagonal entries. 

12. Show that if $X$ and $Y$ are eigenvectors belonging to the same eigenvalue $\lambda$, then $aX + bY$ is also an eigenvector belonging to $\lambda$ (provided $aX + bY \neq 0$). 

13. How are the eigenvalues of $A^T$ related to the eigenvalues of $A$? [**Hint:** Use $(B + C)^T = B^T + C^T$ on the characteristic equation.] 

14. State conditions on $a, b, c, \text{ and } d$ that will guarantee real eigenvalues for 

$$
A = \begin{pmatrix} a & b \\
c & d \end{pmatrix}
$$

15. How are the eigenvalues of $\bar{A}$ related to the eigenvalues of $A$? 

16. How are the eigenvalues of $A^*$ related to the eigenvalues of $A$? 

17. Let $(\lambda, X)$ be an eigenpair of $A$. Show that $(\lambda^n, X)$ is an eigenpair of $A^n$. 

18. Show that if $(\lambda, X)$ is an eigenpair of $A$, then 

$$
\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_2\lambda^2 + a_1\lambda + a_0
$$

is an eigenvalue of 

$$
a_nA^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I
$$
19. Let $A$ be a matrix which is nilpotent. Show that zero is the only eigenvalue of $A$.

5.3 DIAGONALIZATION OF MATRICES

Markov chains are a prime example of the importance of being able to write a matrix $A$ as $PDP^{-1}$, where $D$ is diagonal. When this can be done, we call $A$ diagonalizable.

**Definition 5.3.1.** A matrix $A$ is diagonalizable when there exist a diagonal matrix and an invertible matrix $P$ such that $A = PDP^{-1}$. When $D$ and $P$ are found for a given $A$, we say that $A$ has been diagonalized. Note that $D = P^{-1}AP$.

With the concept of diagonalization defined, we can state another major problem of linear algebra, our fifth.

**Diagonalization Problem**

Given a matrix $A_{n \times n}$, determine whether $A$ is diagonalizable. If $A$ is diagonalizable, find $P$ and $D$ in the equation

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

To approach the diagonalization problem, we first ask: If $A$ is diagonalizable, what must be true about $P$ and $D$? If $A$ is diagonalizable, then $A = PDP^{-1}$ which means that $AP = PD$. Now writing

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & \cdots & p_{nn} \end{pmatrix}$$

and

$$D = \begin{pmatrix} c_1 & 0 & & \\ & c_2 & & \\ & & \ddots & \\ 0 & & & c_n \end{pmatrix}$$

we see that

$$AP_1 = c_1 P_1 \quad AP_2 = c_2 P_2 \quad \ldots \quad AP_n = c_n P_n$$
where \( P_k \) is the vector made of the \( k \)th column of \( P \). Therefore \( P \) is the matrix made up of columns which are eigenvectors of \( A \). The diagonal elements of \( D \) are the corresponding eigenvalues. Moreover, since \( P \) is invertible, the columns are linearly independent. Therefore we have the following theorem.

**Theorem 5.3.1.** If \( A_{n \times n} \) is diagonalizable, then \( A \) has \( n \) linearly independent eigenvectors. Also, in the equation \( A = PD P^{-1} \), \( P \) is a matrix whose columns are eigenvectors, and the diagonal entries of \( D \) are the eigenvalues corresponding column by column to their respective eigenvectors.

This theorem tells us what \( P \) and \( D \) must look like if \( A \) is diagonalizable. We would like this theorem to be reversible. That is, we hope that if we

1. Solve the eigenproblem for \( A_{n \times n} \): \( AX_1 = \lambda_1 X_1, \ldots, AX_n = \lambda_n X_n \),
2. Find that the eigenvectors can be chosen as linearly independent,
3. Set \( P = (X_1 X_2 \cdots X_n) \),

then we would have

\[
A = P \begin{pmatrix}
\lambda_1 & 0 & & \\
& \lambda_2 & & \\
& & \ddots & \\
& & & \lambda_n
\end{pmatrix} P^{-1}
\]

In fact, this is true. Before proving that this procedure works we give an example.

**Example 1.** If possible, diagonalize

\[
A = \begin{pmatrix}
1 & 1 \\
-2 & 4
\end{pmatrix}
\]

**Solution** The eigenproblem for \( A \) was solved in Sec. 5.2. The general eigenpairs are

\[
\begin{pmatrix} 2, \binom{k}{k} \end{pmatrix}, \begin{pmatrix} 3, \binom{j}{2j} \end{pmatrix}
\]

and specific choices are

\[
\begin{pmatrix} 2, \binom{1}{1} \end{pmatrix}, \begin{pmatrix} 3, \binom{1}{2} \end{pmatrix}
\]
5.3. DIAGONALIZATION OF MATRICES

Now
\[ \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\} \]
is a linearly independent set, so we form
\[ P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \]

Then
\[ P^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \]

Finally we check to see whether
\[ A = PDP^{-1} \]

We have
\[ PDP^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \]

\[ = \begin{pmatrix} 2 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \]
\[ = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix} = A \]

Thus \( A \) has been diagonalized. If we had formed
\[ P = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \]

then \( D \) would be
\[ \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \]

Now we state and prove the theorem which the last example illustrates.
Theorem 5.3.2. If $A_{n \times n}$ has $n$ linearly independent eigenvectors $X_1, X_2, \ldots, X_n$ with

$$AX_1 = \lambda_1 X_1 \quad \cdots \quad AX_n = \lambda_n X_n$$

then $A$ is diagonalizable and $A = PDP^{-1}$, where

$$P = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix} \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \lambda_2 & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \lambda_n & \end{pmatrix}$$

Proof. If $\{X_1, \ldots, X_n\}$ is a linearly independent set, then $P = (X_1 \cdots X_n)$ is invertible. Now

$$AP = \begin{pmatrix} AX_1 \\ AX_2 \\ \vdots \\ AX_n \end{pmatrix} = \begin{pmatrix} \lambda_1 X_1 \\ \lambda_2 X_2 \\ \vdots \\ \lambda_n X_n \end{pmatrix} = (X_1 \cdots X_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \lambda_2 & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \lambda_n & \end{pmatrix} = PD$$

Since $AP = PD$, we have $A = PDP^{-1}$ and $A$ has been diagonalized. $\square$

Theorems 5.3.1 and 5.3.2 together give us an important result.

Theorem 5.3.3. Solution to the diagonalization problem

Matrix $A_{n \times n}$ is diagonalizable if and only if $A$ and $n$ linearly independent eigenvectors. In that case, if $X_1, X_2, \ldots, X_n$ are the linearly independent eigenvectors and the eigenpairs are

$$(\lambda_1, X_1), \ (\lambda_2, X_2), \ \cdots, \ (\lambda_n, X_n)$$

then setting $P = (X_1 X_2 \cdots X_n)$ and

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

we have

$$A = PDP^{-1} \quad \text{and} \quad D = P^{-1}AP$$
5.3. **DIAGONALIZATION OF MATRICES**

The result in Theorem 5.3.3 can be stated in two other equivalent ways.

Matrix $A_{n \times n}$ is diagonalizable if and only if there exists a basis of $\mathcal{C}_{n1}$ consisting of eigenvectors of $A$. In that case, if $\{X_1, \ldots, X_n\}$ is the basis of eigenvectors, and the eigenpairs are $(\lambda_1, X_1), \ldots, (\lambda_n, X_n)$, then the construction of $P$ and $D$ proceeds as in the state above.

Let $A_{n \times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_m$ with $m \leq n$ ($m$ is strictly less than $n$ if some of the eigenvalues have multiplicity 2 or more). Then $A$ is diagonalizable if and only if

$$\sum_{k=1}^{m} \dim E_{(\lambda_k)} = n$$

Note that to solve the diagonalization problem for $A$, we first solve the eigenproblem for $A$.

**Example 2.** Solve the diagonalization problem for

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

**Solution** We found in a previous example that the eigenpairs for $A$ are

$$\left( i, c \begin{pmatrix} -i \\ 1 \end{pmatrix} \right), \quad \left( -i, d \begin{pmatrix} i \\ 1 \end{pmatrix} \right)$$

We have two linearly independent eigenvectors for the $2 \times 2$ matrix. Thus $A$ is diagonalizable. Choosing $c = d = 1$, we can select

$$X_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$

and construct

$$P = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$$

Thus

$$P^{-1} \begin{pmatrix} i & 1 \\ -i & 1 \\ 2 & 2 \end{pmatrix}$$
and
\[
\begin{pmatrix}
0 & 1 \\
-1 & 0 \\
\end{pmatrix}
= \begin{pmatrix}
-i & i \\
1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
i & 0 \\
0 & -i \\
\end{pmatrix}
\begin{pmatrix}
i/2 & 1/2 \\
-i/2 & 1/2 \\
\end{pmatrix}
\]

**Example 3.** Solve the diagonalization problem for
\[
A = \begin{pmatrix}
1 & 1 \\
0 & 1 \\
\end{pmatrix}
\]

**Solution** The characteristic equation is
\[
\det \left( \begin{pmatrix}
1 - \lambda & 1 \\
0 & 1 - \lambda \\
\end{pmatrix} \right) = 0
\]
or \((1 - \lambda)^2 = 0\). The eigenvalues are \(\lambda_1 = 1\) and \(\lambda_2 = 1\). For the eigenvectors we solve \(A - \lambda I = 0\) which reduces to
\[
\lambda = 1 \quad \begin{array}{c}
0x_1 + x_2 = 0 \\
0x_1 + 0x_2 = 0
\end{array}
\]

Therefore,
\[
X_1 = \begin{pmatrix}
k \\
0
\end{pmatrix} \quad X_2 = \begin{pmatrix}
j \\
0
\end{pmatrix}
\]

and we cannot obtain two linearly independent eigenvectors for the \(2 \times 2\) matrix. Thus \(A\) is **not** diagonalizable. Another way to say this is that the eigenvectors of \(A\) do not form a basis for \(\mathbb{C}^2\).

**Example 4.** Solve the diagonalization problem for
\[
A = \begin{pmatrix}
-4 & -4 & -8 \\
4 & 6 & 4 \\
6 & 4 & 10
\end{pmatrix}
\]

**Solution** In Example 5 of Sec. 5.2 we solved the eigenproblem for \(A\). Referring to that example, we see that specific eigenpairs are
\[
\begin{pmatrix}
2, \begin{pmatrix}
-2 \\
1
\end{pmatrix}
\end{pmatrix} \quad \begin{pmatrix}
4, \begin{pmatrix}
-1 \\
0
\end{pmatrix}
\end{pmatrix} \quad \begin{pmatrix}
6, \begin{pmatrix}
-2 \\
1
\end{pmatrix}
\end{pmatrix}
\]
5.3. **DIAGONALIZATION OF MATRICES**

Since
\[
\begin{pmatrix}
-2 & 1 & 1 \\
0 & 0 & 1 \\
0 & 1 & 2
\end{pmatrix}
\]
is linearly independent, the 3 \times 3 matrix possesses three linearly independent eigenvectors and is diagonalizable. A choice for \(P\) is
\[
P = \begin{pmatrix}
-2 & -1 & -2 \\
1 & 0 & 1 \\
1 & 1 & 2
\end{pmatrix}
\]
which gives
\[
P^{-1} = \begin{pmatrix}
-1 & 0 & -1 \\
-1 & -2 & 0 \\
1 & 1 & 1
\end{pmatrix}
\]
and
\[
A = P \begin{pmatrix}
2 & 0 & 0 \\
0 & 4 & 0 \\
0 & 0 & 6
\end{pmatrix} P^{-1}
\]
If we wanted \(D\) to have the eigenvalues of \(A\) in descending order of magnitude, we would choose
\[
P = \begin{pmatrix}
-2 & -1 & -2 \\
1 & 0 & 1 \\
2 & 1 & 1
\end{pmatrix}
\]
In that case
\[
D = \begin{pmatrix}
6 & 0 & 0 \\
0 & 4 & 0 \\
0 & 0 & 2
\end{pmatrix} \quad \text{and} \quad P^{-1} = \begin{pmatrix}
1 & 1 & 1 \\
-1 & -2 & 0 \\
-1 & 0 & -1
\end{pmatrix}
\]

**Example 5.** Show that the matrices
\[
A = \begin{pmatrix}
3 & -1 & 0 \\
0 & 3 & 0 \\
0 & 0 & -5
\end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix}
0 & -3 & -3 \\
-3 & 0 & -3 \\
-3 & -3 & 0
\end{pmatrix}
\]
both have 3 as an eigenvalue of multiplicity 2. Show that \(A\) is not diagonalizable, but \(B\) is diagonalizable.
Solution  The characteristic polynomial for $B$ factors as $-(\lambda + 6)(\lambda - 3)^2$, so that 3 is an eigenvalue of multiplicity 2. We already know that 3 is an eigenvalue of multiplicity 2 for $A$ from Example 6 of Sec. 5.2. The eigenpairs for $B$ are

\[
\begin{pmatrix}
3, \\ 
\begin{pmatrix}
-r - s \\
  r \\
 s 
\end{pmatrix}
\end{pmatrix},
\begin{pmatrix}
-6, \\ t \\
 t 
\end{pmatrix}
\]

and we can choose

\[
S = \left\{ \begin{pmatrix}
-1 \\
 1 \\
 0
\end{pmatrix},
\begin{pmatrix}
-1 \\
 0 \\
 1
\end{pmatrix},
\begin{pmatrix}
1 \\
 1 \\
 1
\end{pmatrix} \right\}
\]

as a basis of eigenvectors of $C_{31}$. Thus $B$ is diagonalizable. But the eigenpairs for $A$ do not generate a basis for $C_{31}$, and $A$ is not diagonalizable. The key observation there is that for $A$, dim $E_{(3)}$ is strictly less than the multiplicity of the eigenvalue 3; for $B$, dim $E_{(3)}$ equals the multiplicity of the eigenvalue 3.

Example 6. Determine whether the shear linear transformation as defined in previous examples is diagonalizable.

Solution  To solve this problem, we use a matrix which represents shear. The reason this can be done is that if $A$ and $B$ are similar matrices and one is similar to a diagonal matrix $D$, then the other is also similar to the same diagonal matrix (Probl. 14 in Sec. 4.4). Thus diagonalizability is invariant under similarity, and we say a linear transformation is diagonalizable if some representing matrix of the transformation is diagonalizable. It is sufficient to use the matrix with respect to the standard basis

\[
\begin{pmatrix}
1 & k \\
0 & 1
\end{pmatrix}
\]

to represent shear. The general eigenpair for this matrix is

\[
\begin{pmatrix}
1, \\ r
\end{pmatrix}
\]
5.3. **Diagonalization of Matrices**

A basis for $C_{21}$ cannot be constructed from the eigenvectors of the representing matrix. Therefore, the shear transformation is not diagonalizable.

We now know that an $n \times n$ matrix $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors. If $n$ is large, checking for linear independence can be tedious. There is a simple sufficient condition for diagonalizability.

**Theorem 5.3.4.** If $A_{n\times n}$ has $n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the eigenvectors $X_1, X_2, \ldots, X_n$ in the eigenpairs

$$(\lambda_1, X_1) \quad (\lambda_2, X_2) \quad \ldots \quad (\lambda_n, X_n)$$

form a linearly independent set, and $A$ is therefore diagonalizable.

**Example 7.** Show that

$$A = \begin{pmatrix} 1 & 7 & 6 \\ 0 & -1 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

is diagonalizable.

**Solution** The characteristic equation is $\det(A - \lambda I) = 0$ which is $(1 - \lambda)(-1 - \lambda)(2 - \lambda) = 0$. Thus the eigenvalues of $A$ are $1, -1, \text{ and } 2$. Since $A$ is $3 \times 3$ and $A$ has three distinct eigenvalues, Theorem 5.3.4 implies that $A$ is diagonalizable.

**Proof of Theorem 5.3.4.** Suppose $A_{n\times n}$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, with associated eigenvectors $X_1, X_2, \ldots, X_n$. If we show that $S = \{X_1, \ldots, X_n\}$ is a linearly independent set, then $A$ is diagonalizable. We will suppose that $S$ is linearly dependent and derive a contradiction.

Suppose that $S$ is linearly dependent (LD). We can find a set $S_k = \{X_1, \ldots, X_k\}$, $k < n$, which is linearly independent (LI) by the following process. Since eigenvectors are nonzero, $\{X_1\}$ is an LI set. If $\{X_1, X_2\}$ is LD, we stop with $S_1$ as our LI set. If $S_2$ is LI but $S_3 = \{X_1, X_2, X_3\}$ is LD, we stop with $S_2$, and so on. Suppose we have completed this process and have $S_k$ LI but $S_{k+1}$ LD ($k + 1$ could equal $n$).

So there exist constants $c_1, \ldots, c_{k+1}$, not all zero, with

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k + c_{k+1}X_{k+1} = 0 \quad (5.3.1)$$
Also $c_{k+1} \neq 0$, because if $c_k$ is zero, then $c_1 = c_2 = \cdots = c_k = 0$ (by the linear independence of $S_k$), and $S_{k+1}$ would be LI, a contradiction.

Now we multiply both sides of Eq. (5.3.1) and use matrix algebra to obtain

$$c_1AX_1 + c_2AX_2 + \cdots + c_kAX_k + c_{k+1}AX_{k+1} = 0$$

or

$$c_1\lambda_1X_1 + c_2\lambda_2X_2 + \cdots + c_k\lambda_kX_k + c_{k+1}\lambda_{k+1}X_{k+1} = 0 \quad (5.3.2)$$

Multiplying (5.3.1) by $\lambda_{k+1}$ and subtracting from Eq. (5.3.2), we have

$$c_1(\lambda_1 - \lambda_{k+1})X_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})X_k = 0$$

Since the eigenvalues are distinct and $S_k$ is LI, we must have $c_1 = c_2 = \cdots = c_k = 0$. Substituting these values in (5.3.1), we have

$$c_{k+1}X_{k+1} = 0$$

Now $c_{k+1} \neq 0$ so $X_{k+1}$ must be the zero vector; this contradicts the hypothesis that $X_{k+1}$ is an eigenvector. Therefore it is impossible for $S$ to be LD, and $S$ must be LI. ∎

**Checking Diagonalizations of Matrices**  Once a matrix has been diagonalized, we may want some easy checks on the diagonal form $D$. Since

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

means that $A$ is similar to $D$, then by previous results on similar matrices we must have

$$\text{tr } t = \text{tr } D$$

$$\det A = \det D$$

$$\text{rank } A = \text{rank } D$$

**Example 8.** Use the three checks above to check the diagonalization in Example 4.

**Solution**  In Example 4 we found that

$$A = \begin{pmatrix} -4 & -4 & -8 \\ 4 & 6 & 4 \\ 6 & 4 & 10 \end{pmatrix} \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$
5.3. **DIAGONALIZATION OF MATRICES**

In this case \( \text{tr } A = \text{tr } D = 12 \), \( \det A = \det B = 48 \), and \( \text{rank } A = \text{rank } D = 3 \). This does not guarantee that the diagonalization is correct but gives more confidence in the answer.

**Example 9.** A person claims that the matrix

\[
A = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}
\]

is similar to

\[
D = \begin{pmatrix}
15 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

Is this correct?

**Solution** We have

\[
\text{tr } A = 15 = \text{tr } D \\
\det A = 0 = \det D \\
\text{rank } A = 2 \neq 1 = \text{rank } D
\]

Since the ranks are unequal, the person has incorrectly diagonalized \( A \).

Note that these checks are not sufficient to prove a diagonalization correct; they can only help you find an incorrect diagonalization. For example,

\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]

have the same trace, determinant, and rank, but

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]

is not the diagonalization of

\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}
\]

because

\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}
\]
is not even diagonalizable. See Example 3.

**Geometric Meaning of Diagonalization** In the real vector space case, the diagonalization of \( A_{n \times n} \) gives information about the geometric action of the transformation \( T_A: \mathcal{M}_{n1} \to \mathcal{M}_{n1} \) generated by \( A \). For example, since

\[
\begin{pmatrix}
1 & 1 \\
-2 & 4
\end{pmatrix}
= \begin{pmatrix}
1 & 1 \\
1 & 2
\end{pmatrix}
\begin{pmatrix}
2 & 0 \\
0 & 3
\end{pmatrix}
\begin{pmatrix}
1 & 1 \\
-1 & 1
\end{pmatrix}
= \begin{pmatrix}
1 & 1 \\
-2 & 4
\end{pmatrix}
\begin{pmatrix}
1 \\
-1
\end{pmatrix}
= 2 \begin{pmatrix}
1 \\
1
\end{pmatrix}
\]

we know that \( T_A \) simply stretches the vectors

\[
\begin{pmatrix}
1 \\
1
\end{pmatrix}
\text{ and } \begin{pmatrix}
1 \\
2
\end{pmatrix}
\]

(See Fig. 5.3.1.) Since

\[
\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}
\]

is a basis for \( E^2 \), if a vector \( X \) has coordinate vector

\[
\begin{pmatrix}
a \\
b
\end{pmatrix}
\]

then the coordinate vector of

\[
T_A\left( \begin{pmatrix} a \\ b \end{pmatrix} \right) \text{ is } \begin{pmatrix} 2a \\ 3b \end{pmatrix}
\]

This illustrates the use of the words **characteristic values and characteristic vectors** in that the vectors

\[
\begin{pmatrix}
1 \\
1
\end{pmatrix}
\text{ and } \begin{pmatrix}
1 \\
2
\end{pmatrix}
\]

are “characteristic” of the action of \( T \). The vectors define directions which can be called **principal directions** of \( T \).

**Example 10.** Show that if the coordinate vector of \( X \) with respect to

\[
S = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}
\]

is \( \begin{pmatrix} a \\ b \end{pmatrix} \)

then

\[
(T_A(X))_S = \begin{pmatrix} 2a \\ 3b \end{pmatrix}
\]
5.3. **DIAGONALIZATION OF MATRICES**

**Solution** For a given vector \( X \), the coordinates are given by \( PX \), where \( P \) is the transition matrix from \( \mathcal{M}_{21} \) with the standard basis to \( \mathcal{M}_{21} \) with basis \( S \). In this case the matrix is

\[
P = \begin{pmatrix}
2 & -1 \\
-1 & 1
\end{pmatrix}
\]

Now

\[
AX = PDP^{-1}X = PDP^{-1}X = PD \begin{pmatrix} a \\ b \end{pmatrix} = P \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = P \begin{pmatrix} 2a \\ 3b \end{pmatrix}
\]

Since

\[
P \begin{pmatrix} 2a \\ 3b \end{pmatrix}
\]

gives the vector \( AX \), and \( P \) is the inverse of the transition matrix \( P^{-1} \),

\[
\begin{pmatrix} 2a \\ 3b \end{pmatrix}
\]

is the coordinate vector with respect to \( S \) for \( AX \).

The last example illustrates the following fact.

If \( A = A_{n \times n} = PD P^{-1} \), then the action of the transformation \( T_A \) on \( \mathcal{M}_{n1} \) can be thought of in terms of the action of \( D \) on \( \mathcal{M}_{n1} \) with the basis of eigenvectors of \( A \).

We will see in the next section that if \( A \) is a real symmetric matrix, then the basis of eigenvectors can always be chosen as orthonormal.

**Example 11.** Analyze \( T: \mathcal{M}_{21} \to \mathcal{M}_{21} \) defined by

\[
T(X) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} X \quad X \in \mathcal{M}_{21}
\]

by diagonalizing the matrix.

**Solution** Let

\[
A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}
\]
and calculate \( \det(A - \lambda I) \). The characteristic equation \( \lambda^2 - \lambda = 0 \) yields eigenvalues \( \lambda_1 = 1 \) and \( \lambda_2 = 0 \). The eigenpairs are

\[
\begin{pmatrix} 1, \begin{pmatrix} k \\ k \end{pmatrix} \\ 0, \begin{pmatrix} -j \\ j \end{pmatrix} \end{pmatrix}
\]

Choosing \( k \) and \( j \) so that the eigenvectors have length 1, we have

\[
\begin{pmatrix} 1, \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix} \\ 0, \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \end{pmatrix}
\]

Therefore

\[
\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ \sqrt{2} & \sqrt{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \sqrt{2} & \sqrt{2} \\ -1 & 1 \end{pmatrix}
\]

This means that the action of \( T \) on a standard coordinate matrix for a vector is as follows:

First: Rotation 45° clockwise
Second: Projection on \( x \) axis
Third: Rotation 45° counterclockwise

As a final application of these remarks, we note that since the shear transformation is not diagonalizable (Example 6), the shear transformation does not “stretch” objects in two independent directions. This reflects our intuitive feelings about shear, which result from a “sideways” deformation of the cube illustrated in previous encounters with this example.

**PROBLEMS 5.3**

1. Solve the diagonalization problem for the following matrices. If the matrix is diagonalizable, write \( P, P^{-1}, \) and \( D \) in \( A = PDP^{-1} \).
5.3. **DIAGONALIZATION OF MATRICES**

\[
\begin{align*}
\text{(a)} & \quad \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} & \quad \text{(b)} & \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{2}{3} \end{pmatrix} & \quad \text{(c)} & \quad \begin{pmatrix} -2 & -2 & -4 \\ 2 & 3 & 2 \\ 3 & 2 & 5 \end{pmatrix} \\
\text{(d)} & \quad \begin{pmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{pmatrix} & \quad \text{(e)} & \quad \begin{pmatrix} 1 & 1 & 0 & 0 \\ -2 & 4 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} & \quad \text{(f)} & \quad \begin{pmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{pmatrix} \\
\text{(g)} & \quad \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} & \quad \text{(h)} & \quad \begin{pmatrix} 1 & 2 \\ -2 & 2 \end{pmatrix} & \quad \text{(i)} & \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \\
\text{(j)} & \quad \begin{pmatrix} i & 1 \\ 0 & i \end{pmatrix} & \quad \text{(k)} & \quad \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}
\end{align*}
\]

2. Show that if a diagonalization of \( A \) is \( D \), then \( D^2 \) is a diagonalization of \( A^2 \). What is a diagonalization of \( A^n \)?

3. Diagonalize

\[
A = \begin{pmatrix} 1 & 1 \\ 0 & \frac{1}{2} \end{pmatrix}
\]

and use the diagonalization to calculate \( A^{12} \).

4. If \( A \) is diagonalizable, is \( A^T \) diagonalizable?

5. If \( A \) is nonsingular and diagonalizable, is \( A^{-1} \) diagonalizable?

6. (a) If \( A \) is diagonalizable and trace \( A > 0 \), must \( A \) have a positive eigenvalue?

(b) What if \( A \) is required to have real entries and all real eigenvalues?

7. If \( A \) is diagonalizable, has all positive entries, and has all real eigenvalues, must \( A \) have a positive eigenvalue?

8. Let \( A \) be a nilpotent matrix, with \( A^m = 0 \). If \( A \) is diagonalizable, what are its eigenvalues?

9. Let \( A \) be an involutory matrix \( (A^2 = I) \). If \( A \) is diagonalizable, what are its eigenvalues?

10. Let \( A \) be an orthogonal matrix \( (AA^T = I) \). If \( A \) is diagonalizable, what can you say about its eigenvalues?
11. If $A$ and $B$ are diagonalizable with

\[ A = P^{-1}D_1 P \quad B = Q^{-1}D_2 Q \]

is $D_1 D_2$ necessarily a diagonalization of $AB$? Is $D_1 + D_2$ necessarily a diagonalization of $A + B$?

12. Analyze the action of $T: \mathcal{M}_{21} \to \mathcal{M}_{21}$ defined by

\[ T(X) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} X \quad X \in \mathcal{M}_{21} \]

by diagonalizing the matrix.

5.4 DIAGONALIZATION: SYMMETRIC AND HERMITIAN MATRICES

Symmetric and hermitian matrices, which arise in many applications, enjoy the property of always being diagonalizable. Also the set of eigenvectors of such matrices can always be chosen as orthonormal. The diagonalization procedure is essentially the same as outlined in Sec. 5.3, as we will see in our examples.

**Example 1.** The horizontal motion of the system of masses and springs where all the masses are the same and the springs are the same, can be analyzed by diagonalizing the symmetric matrix

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

Diagonalize $A$.

**Solution** We have

\[ \det(A - \lambda I) = \det \begin{pmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{pmatrix} = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 \]

so that the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 1$. Eigenvectors are found by solving $(A - \lambda I)X = 0$; these equations and solutions are

$\lambda_1 = 3$: $-1x_1 - 1x_2 = 0$ \Rightarrow $X_1 = \begin{pmatrix} k \\ -k \end{pmatrix}$
5.4. **DIAGONALIZATION**

\[ \lambda_2 = 1: \begin{align*}
1x_1 - 1x_2 &= 0 \\
-1x_1 - 1x_2 &= 0 \Rightarrow X_2 = \begin{pmatrix} j \\ j \end{pmatrix}
\end{align*} \]

Now \( X_1 \) and \( X_2 \) are orthogonal since \( X_1 \cdot X_2 = k_j - k_j = 0 \). If we normalize \( X_1 \) and \( X_2 \), we have the choices for eigenvectors

\[ O_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}, \quad O_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \]

and \( S = \{O_1, O_2\} \) forms an orthonormal set. Finally, with

\[ P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \]

we can write

\[ \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \]

Since \( P \) is an orthogonal matrix, \( P^{-1} = P^T \). Also \( P \) represents a rotation of \( \pi/4 \) radians clockwise.

**Example 2.** Diagonalize

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

Choose \( P \) as an orthogonal matrix.

**Solution** The characteristic equation is \((1 - \lambda)(\lambda^2 - 1) = 0\), which has solutions 1 and \(-1\). The eigenvalue 1 has multiplicity 2, and the eigenvalue \(-1\) is simple (multiplicity 1). Now to determine the eigenvectors, we solve equations.

\[ \lambda_1 = 1: \begin{align*}
-x_1 + x_2 + 0x_3 &= 0 \\
x_1 - x_2 + 0x_3 &= 0 \Rightarrow X_1 = \begin{pmatrix} j \\ j \\ k \end{pmatrix} \\
0x_1 + 0x_2 + 0x_3 &= 0
\end{align*} \]

\[ \lambda_2 = -1: \begin{align*}
1x_1 + 1x_2 + 0x_3 &= 0 \\
1x_1 + 1x_2 + 1x_3 &= 0 \Rightarrow X_2 = \begin{pmatrix} r \\ -r \\ 0 \end{pmatrix} \\
0x_1 + 0x_2 + 2x_3 &= 0
\end{align*} \]
We see that corresponding to $\lambda_1 = 1$ the eigenspace is two-dimensional. Therefore, we can choose a basis for $E_{\lambda_1}$ (the basis problem has returned!) of orthonormal vectors. In fact, choosing first $j = 1, k = 0$ and second $k = 1, j = 0$, we have
\[ V_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad V_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \]
which are orthogonal. Normalizing $V_1$ ($V_2$ is already normalized) yields
\[ O_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \quad \text{and} \quad O_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \]
Not as much work is required for $\lambda_2 = -1$ since the eigenspace for this eigenvalue is one-dimensional. Normalizing $X_2$, we have
\[ O_3 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \]
Therefore
\[ P = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad P^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix} = P^T \quad (!!) \]
Diagonalize $A$:
\[ A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} P^T \]

The last two examples illustrate the basic results for diagonalization of symmetric matrices.

**Theorem 5.4.1.** If $A_{n \times n}$ is symmetric with real entries, then

(a) The eigenvalues are real.

(b) Eigenvectors corresponding to distinct eigenvalues are orthogonal.
5.4. **DIAGONALIZATION**

(c) The eigenspaces of each eigenvalue have orthogonal bases. The dimension of an eigenspace corresponds to the multiplicity of the eigenvalue.

(d) Matrix $A$ is orthogonally diagonalizable; that is, there exists an orthogonal matrix $P$ such that

$$A = PDP^T \quad (\text{and so} \quad D = P^TAP)$$

**Proof.** We prove only parts (a) and (b). Parts (c) and (d) are proved in more advanced texts.

(a) Suppose that $\lambda = a + bi$ and that $X + iY$ is the corresponding eigenvector. Therefore,

$$[A - (a + bi)I](X + iY) = 0 + 0i$$

Carrying out the multiplications and setting real and imaginary parts equal, we find

$$aIX - AX - bIY = 0$$
$$aIY - AY + bIX = 0$$

Now take the dot product of the first equation with $Y$ and the second with $X$ to find

$$aY \cdot IX - Y \cdot AX - bY \cdot IY = Y \cdot 0 = 0$$
$$aX \cdot IY - X \cdot AY + bY \cdot IX = X \cdot 0 = 0$$

or

$$aY^TIX - Y^TAX - bY^T IY = 0$$
$$aX^TIY - X^TAY + bX^TIX = 0$$

Because $A$ and $I$ are symmetric $I = I^T$ and $A = A^T$. Thus

$$aY^TI^TX^T - Y^TA^TX^T - bY^T IY = 0$$
$$aX^TIY - X^TAY + bX^TIX = 0$$

and since the transpose of a product is the product of the transposes in reverse order, we have

$$a(X^TIY)^T - (X^T AY)^T - bY^T IY = 0$$
$$aX^TIY - X^T AY + bX^TIX = 0$$
Finally, taking the transpose of both sides of the second equation results in
\[
\begin{align*}
    a(X^T IY)^T - (X^T AY)^T & = 0 \\
    a(X^T IY)^T - (X^T AY)^T + bX^T I X & = 0
\end{align*}
\]
Subtracting the equations yields
\[
b(Y^T IY + X^T I X) = 0
\]
or
\[
b(|Y|^2 + |X|^2) = 0
\]
Because both |X| and |Y| cannot be zero (for if so, \(X + iY = 0\) and could not be an eigenvector), we must have \(b = 0\). Therefore \(\lambda\) is real.

(b) Let \(X_1\) and \(X_2\) be eigenvectors corresponding to \(\lambda_1\) and \(\lambda_2\), \(\lambda_1 \neq \lambda_2\), \(\lambda_1 \neq 0\). We want to show that \(X_1 \cdot X_2 = 0\). Now
\[
X_1 \cdot X_2 = \frac{1}{\lambda_1} \lambda_1 X_1 \cdot X_2 = \frac{1}{\lambda_1} AX_1 \cdot X_2
\]
\[
= \frac{1}{\lambda_1} (AX_1)^T X_2
\]
\[
= \frac{1}{\lambda_1} (X_1^T A^T) X_2
\]
(By symmetry)
\[
= \frac{1}{\lambda_1} (X_1^T A) X_2
\]
\[
= \frac{1}{\lambda_1} X_1^T (AX_2)
\]
\[
= \frac{1}{\lambda_1} X_1^T \lambda_2 X_2
\]
\[
= \frac{\lambda_2}{\lambda_1} X_1^T X_2
\]
\[
= \frac{\lambda_2}{\lambda_1} (X_1 \cdot X_2)
\]
If \(\lambda_2 = 0\), then \(X_1 \cdot X_2 = 0\). If \(\lambda_2 \neq 0\), then \(\lambda_2/\lambda_1 \neq 1\) and we must still have \(X_1 \cdot X_2 = 0\).

Theorem 5.4.1c tells us that we can find an orthogonal basis for each eigenspace. This may require the Gram-Schmidt process, as the next example shows.
Example 3. Orthogonally diagonalize

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

That is, diagonalize \( A \) with an orthogonal matrix \( P \).

Solution  The characteristic polynomial is \(-\lambda^3 + 3\lambda + 2\) which has roots \(-1\) (multiplicity 2) and 2 (simple). To determine eigenvectors, we solve \((A - \lambda I)X = 0\):

\[
\begin{align*}
\lambda &= 2: \\
-2x_1 + x_2 + x_3 &= 0 \\
x_1 - 2x_2 + x_3 &= 0 \Rightarrow X_1 = \begin{pmatrix} k \\ k \\ k \end{pmatrix} \\
x_1 + x_2 - 2x_3 &= 0
\end{align*}
\]

\[
\begin{align*}
\lambda &= 2: \\
x_1 + x_2 + x_3 &= 0 \\
x_1 + x_2 + x_3 &= 0 \Rightarrow X_2 = \begin{pmatrix} -k - j \\ k \\ j \end{pmatrix}
\end{align*}
\]

Since rank \((A - \lambda_2 I) = 1\), the dimension of \(E_{(\lambda_2)}\) is 2.
Looking at \(X_2\) and putting \(k = 1, j = 0\), we have

\[ V_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \]

in the eigenspace. Setting \(k = 0, j = 1\), we find

\[ V_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \]

in the eigenspace. Now \(V_1\) and \(V_2\) are not orthogonal to each other, but they are linearly independent and span the eigenspace. Using the Gram-Schmidt process on \(\{V_1, V_2\}\), we find

\[ O_1 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \quad \text{and} \quad O_2 = \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix} \]
as orthonormal basis vectors for the eigenspace of \( \lambda_2 = -1 \). Letting \( O_3 = X_1/|X_1| \), we obtain an orthonormal basis (for \( E^3 \)) of eigenvectors of \( A \). Choosing \( P \) as

\[
\begin{pmatrix}
-1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\
1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\
0 & 2/\sqrt{6} & 1/\sqrt{3}
\end{pmatrix}
\]

we have

\[ A = P \begin{pmatrix}
-1 & 0 & 0 \\
0 & -1 & 0 \\
0 & 0 & 2
\end{pmatrix} P^T. \]

Now that we can orthogonally diagonalize symmetric matrices, we can consider an application to analytic geometry.

**Quadratic Forms and Conic Sections** A classical problem of analytic geometry is the following:

For the conic section centered at the origin of the \( xy \) plane, described by

\[ ax^2 + bxy + cy^2 = d \quad (5.4.1) \]

determine whether the conic section is an ellipse, a hyperbola, or a parabola. Graph the conic section.

A symmetric matrix can be used to describe the left-hand side of Eq. 5.4.1. In particular,

\[ \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + bxy + cy^2 = d \]

Let us call

\[ A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \]

the **matrix of the conic section**. Making a change of basis with the orthogonal matrix \( P \) which diagonalizes \( A \), we write

\[ \begin{pmatrix} x'' \\ y'' \end{pmatrix} = P^T \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{or} \quad P \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \]

\[ ^3\text{Several matrices are possible} \]

\[ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \]

are two. We used the symmetric matrix so we can be assured of diagonalizability.
5.4. **DIAGONALIZATION**

Substituting and denoting
\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix}
\]
by \( X' \), we have
\[
X^TAX = \begin{pmatrix} X' \end{pmatrix}^T A \begin{pmatrix} X' \end{pmatrix} = \begin{pmatrix} X' \end{pmatrix}^T \begin{pmatrix} P^T \end{pmatrix} A \begin{pmatrix} P \end{pmatrix} X' = X'^T DX'
\]
which reduces to
\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix}^T \begin{pmatrix}
  \lambda_1 & 0 \\
  0 & \lambda_2
\end{pmatrix} \begin{pmatrix}
  x' \\
  y'
\end{pmatrix} = d
\]
or
\[
\lambda_1 x'^2 + \lambda_2 y'^2 = d
\]
The last equation is easy to classify and graph in the \( x'y' \) plane since it has no “mixed” term \( x'y' \).

**Example 4.** Classify \( xy = 1 \) and graph it.

**Solution** (We already know the graph since the equation can be rewritten as \( y = 1/x \). This will make it easy to check our answer.) The matrix of the conic section is
\[
A = \begin{pmatrix}
  0 & \frac{1}{2} \\
  \frac{1}{2} & 0
\end{pmatrix}
\]
and \( A \) has eigenpairs
\[
\begin{pmatrix}
  \frac{1}{2} \\
  \frac{1}{2}
\end{pmatrix}, \begin{pmatrix}
  1/\sqrt{2} \\
  1/\sqrt{2}
\end{pmatrix}
\]
so that
\[
P = \begin{pmatrix}
  1/\sqrt{2} & -1/\sqrt{2} \\
  1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix}, \quad p^T = \begin{pmatrix}
  1/\sqrt{2} & 1/\sqrt{2} \\
  -1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix}
\]
and
\[
A = P \begin{pmatrix}
  \frac{1}{2} & 0 \\
  0 & \frac{1}{2}
\end{pmatrix} p^T
\]
Therefore, with the change of variables (basis)
\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix} = P^T \begin{pmatrix}
  x \\
  y
\end{pmatrix} = \begin{pmatrix}
  1/\sqrt{2} & 1/\sqrt{2} \\
  -1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix} \begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]
which is the same as

\[ x' = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y \quad y' = -\frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y \]

we have

\[ \frac{1}{2}x'^2 - \frac{1}{2}y'^2 = 1 \]

or

\[ \frac{x'^2}{(\sqrt{2})^2} - \frac{y'^2}{(\sqrt{2})^2} = 1 \]

Therefore, the conic section is a hyperbola. To sketch the graph, we must determine the \( x', y' \) axes. Since the point \((x, y) = (1, 1)\) gives \((x', y') = (2/\sqrt{2}, 0)\) and \((x, y) = (-1, 1)\) gives \((x', y') = (0, 2/\sqrt{2})\), the \( x' \) and \( y' \) axes are as shown in Fig. 5.4.1. The graph of the hyperbola is also shown in Fig. 5.4.1. Note that the new axes contain the eigenvectors.

**Example 5.** Classify the conic section

\[ 2x^2 + 2xy + 2y^2 = 27 \]

and graph it.

**Solution** The matrix of the conic section is

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

Eigenpairs of \( A \) are

\( \left( 3, \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right) \) and \( \left( 1, \left( -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right) \)

Therefore in \( x', y' \) coordinates defined as in Example 4 (the eigenvectors are the same) we have

\[ 3x'^2 + y'^2 = 27 \]

or

\[ \frac{x'^2}{3^2} + \frac{y'^2}{(3\sqrt{3})^2} = 1 \]

which describes an ellipse. The graph of the ellipse is shown in Fig. 5.4.2. Note that the new axes contain the eigenvectors of the matrix. Also note
that \( x'y' \) axes are obtained by a 45° counterclockwise rotation, which is the action of \( P \). Moreover, \( x' \) is defined by the first eigenvector, and \( y' \) is defined by the second eigenvector.

Those who have solved these types of conic section problems in calculus realize that this linear algebra method of removing the \( xy \) term is much simpler. Of course, a lot of power machinery had to be developed to get to this point.

In general, the problem of removing the \( xy \) term in \( ax^2 + bxy + cy^2 \) is known as the problem of **diagonalizing a quadratic form**. This problem arises in many areas; statistics and physics are two. A real quadratic form in the variables \( x_1, x_2, \ldots, x_n \) is a function \( Q : \mathbb{R}^n \to \mathbb{R} \) given by

\[
Q \left( \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right) = (x_1 \ x_2 \ \cdots \ x_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{5.4.2}
\]

where \( A \) is in \( \mathcal{M}_{nn} \). Written out, a real quadratic form in \( x_1, x_2, x_3 \) looks like

\[
a x_1^2 + b x_2^2 + c x_3^2 + d x_1 x_2 + e x_1 x_3 + f x_2 x_3
\]

where \( a \) through \( f \) are real numbers. Note that each term has degree 2—hence the name **quadratic form**. Our basic theorem about diagonalization of symmetric matrices means that any real quadratic form can be diagonalized. So there are new variables \( x'_1, \ldots, x'_n \) such that in the new variables \( X' = P^TX \)

\[
Q \left( \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} \right) = \lambda_1 x'_1^2 + \lambda_2 x'_2^2 + \cdots + \lambda_n x'_n^2
\]

This follows from the fact that the matrix \( A \) in Eq. (5.4.2) can always be chosen as symmetric, and symmetric matrices are orthogonally diagonalizable.

**Diagonalization in the Hermitian Case** Theorem 5.4.1 with a slight change of wording holds true for hermitian matrices.

If \( A_{n \times n} \) is hermitian, then

1. The eigenvalues are real.
2. Eigenvectors corresponding to distinct eigenvalues are orthogonal.

3. The eigenspaces of each eigenvalue have orthogonal bases. The dimension of an eigenspace corresponds to the multiplicity of the eigenvalue.

4. Matrix $A$ is **unitarily diagonalizable**. That is, there exists a unitary matrix $U(U^{-1} = U^*)$ such that

$$A = UDU^* \quad \text{(thus } D = U^*AU)$$

The proofs of 1 and 2 are almost the same as in Theorem 5.4.1a and b. The difference is that $A^*$ is used instead of $A^T$ and in $C_n$, $X \cdot Y = X^*Y$.

**Example 6.** Can

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

be unitarily diagonalized? If so, perform the diagonalization.

**Solution**

$$A^* = A^T = \begin{pmatrix} 1 & 1 - i \\ 1 + i & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 1 - i \\ 1 + i & 0 \end{pmatrix} = A$$

Because $A$ is hermitian, it can be unitarily diagonalized. Now to find the eigenpairs,

$$\det(A - \lambda I) = (1 - \lambda)(-\lambda) - (1 - i)(1 + i) = \lambda^2 - \lambda - 2$$

So we have $\lambda_1 = 2$ and $\lambda_2 = -1$. Eigenpairs are

$$\left(2, \begin{pmatrix} 1 - i \\ 1 \end{pmatrix}\right), \left(-1, \begin{pmatrix} 1 - i \\ -2 \end{pmatrix}\right)$$

To find $U$, we normalize the eigenvectors and use them for the columns of $U$. The normalized eigenvectors are found by calculating

$$\left\| \begin{pmatrix} 1 - i \\ 1 \end{pmatrix} \right\| = \sqrt{(1 + i, 1) \begin{pmatrix} 1 - i \\ 1 \end{pmatrix}} = \sqrt{3}$$

and

$$\left\| \begin{pmatrix} 1 - i \\ -2 \end{pmatrix} \right\| = \sqrt{(1 + i, -2) \begin{pmatrix} 1 - i \\ -2 \end{pmatrix}} = \sqrt{6}$$
5.4. DIAGONALIZATION

(Remember that \(|X| = \sqrt{\langle X, X \rangle} = \sqrt{X^*X}\).) Thus we have

\[
U = \begin{pmatrix}
    1 - i & 1 - i \\
    \sqrt{3} & \sqrt{6} \\
    1 & -2 \\
    \sqrt{3} & \sqrt{6}
\end{pmatrix} \quad U^* = \begin{pmatrix}
    1 + i & 1 \\
    \sqrt{3} & \sqrt{3} \\
    1 + i & -2 \\
    \sqrt{6} & \sqrt{6}
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
    1 & 1 - i \\
    1 + i & 0
\end{pmatrix} = U \begin{pmatrix}
    2 & 0 \\
    0 & -1
\end{pmatrix} U^*
\]

When a hermitian matrix \(A\) is diagonalized, the set of orthonormal eigenvectors of \(A\) is called the set of principal axes of \(A\) and the associated matrix \(U\) is called a principal axis transformation. For a real hermitian matrix, the principal axis transformation allows us to analyze \(A\) geometrically.

**Example 7.** Consider \(T_A : \mathcal{M}_{21} \rightarrow \mathcal{M}_{21}\) defined by

\[
T_A \begin{pmatrix}
    a \\
    b
\end{pmatrix} = \begin{pmatrix}
    1 & 2 \\
    2 & 1
\end{pmatrix} \begin{pmatrix}
    a \\
    b
\end{pmatrix}
\]

This can be diagonalized with

\[
U = \begin{pmatrix}
    1/\sqrt{2} & -1/\sqrt{2} \\
    1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix}
\]

so that

\[
\begin{pmatrix}
    1 & 2 \\
    2 & 1
\end{pmatrix} = \begin{pmatrix}
    1/\sqrt{2} & -1/\sqrt{2} \\
    1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix} \begin{pmatrix}
    3 & 0 \\
    0 & -1
\end{pmatrix} \begin{pmatrix}
    1/\sqrt{2} & 1/\sqrt{2} \\
    -1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix}
\]

Now \(U\) represents rotation of 45° counterclockwise while \(U^T\) represents rotation of 45° clockwise. If we want to see what \(T_A\) does to

\[
\begin{pmatrix}
    a \\
    b
\end{pmatrix}
\]

we can look at

\[
\begin{pmatrix}
    1/\sqrt{2} & -1/\sqrt{2} \\
    1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix} \begin{pmatrix}
    3 & 0 \\
    0 & -1
\end{pmatrix} \begin{pmatrix}
    1/\sqrt{2} & 1/\sqrt{2} \\
    -1/\sqrt{2} & 1/\sqrt{2}
\end{pmatrix} \begin{pmatrix}
    a \\
    b
\end{pmatrix}
\]

and we see that \(T_A\) is a
CHAPTER 5. EIGENVALUES AND DIAGONALIZATION

1. Rotation of 45° clockwise

2. Stretch of 3 in the first component and a reversal in the second component

3. Rotation of 45° counterclockwise

We can say even more by determining what $T_A$ does to the unit circle. In the new coordinates, the unit circle is unchanged because $U$ and $U^T$ represent rotations. However, in the new coordinates we have the action of $D$ as changing the unit circle by reflecting about the line which is defined by

$$\text{span} \left\{ \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right\}$$

and then transforming the reflected circle to an ellipse, as shown in Fig. 5.4.3.

Finally, we note that in diagonalizing a quadratic form for a conic section, the new axes obtained from the rotation are exactly the principal axes of the matrix for the quadratic form.

PROBLEMS 5.4

Diagonalize the matrices in Probs. 1 to 7. If possible, orthogonally diagonalize.

1. $\begin{pmatrix} -1 & 2 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & -1 \end{pmatrix}$
2. $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$
3. $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
4. $\begin{pmatrix} 0 & 0 & -2 \\ 0 & -4 & 0 \\ -2 & 0 & 3 \end{pmatrix}$
5. $\begin{pmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$
6. $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$
7. $\begin{pmatrix} 10 & 8 & 4 \\ 8 & 10 & 4 \\ 4 & 4 & 4 \end{pmatrix}$

8. Classify and graph the following conic sections.
   (a) $4x^2 + 4xy + 4y^2 = 18$
   (b) $3x^2 + 10xy + 3y^2 = 8$
9. Show that the rotation matrix

\[ R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \theta \neq 0, \pi \]

is not diagonalizable if complex eigenvalues are not allowed. What if no restriction is placed on eigenvalues?

10. If an antisymmetric matrix is diagonalizable, must one of the eigenvalues be zero?

11. Unitarily diagonalize

\[ \begin{pmatrix} 2 & 3 + 3i \\ 3 - 3i & 5 \end{pmatrix} \]

12. Unitarily diagonalize

\[ \begin{pmatrix} 1 & i \\ -i & 0 \end{pmatrix} \]

13. Unitarily diagonalize

\[ \begin{pmatrix} 1 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \]

14. Unitarily diagonalize

\[ \begin{pmatrix} 2 & 0 & 1 + i \\ 0 & 2 & i \\ 1 - i & -i & 0 \end{pmatrix} \]

15. A matrix is called a Hadamard matrix if its entries are only 1 or -1 and its rows are mutually orthogonal. Show that

\[ \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \]

are Hadamard matrices. (Hadamard matrices are used in coding theory.)
16. If \( H_{n \times n} \) is a Hadamard matrix, must the \( 2n \times 2n \) matrix
\[
\begin{pmatrix}
H & H \\
H & -H
\end{pmatrix}
\]
be a Hadamard matrix?

17. Let \( A \) be real symmetric. Show that \( \det A \) equals to product of the eigenvalues of \( A \).

18. Let \( A \) be hermitian. Show that \( \det A \) is real.

19. Let \( A \) be symmetric with all eigenvalues positive. Show that there is a matrix \( S \) such that \( A = S^2 \). Matrix \( S \) is called the square root of \( A \). [Hint: Use \( A = PDP^T \) and ask yourself, Does \( D \) have a square root?]

20. Use the diagonalization of
\[
A = \begin{pmatrix}
2 & 1 \\
1 & 2
\end{pmatrix}
\]
(see Example 5) to analyze the transformation of \( \mathbb{R}^2 \) defined by \( A \).

5.5 POSTCALCULUS: SYSTEMS OF LINEAR FIRST-ORDER DIFFERENTIAL EQUATIONS, FUNDAMENTAL FREQUENCIES, AND STABILITY

A linear homogeneous first-order system of differential equations, with constant coefficients, in the three functions \( y_1(t), y_2(t), y_3(t) \) is a set of equations of the form (the prime denotes differentiation)
\[
\begin{pmatrix}
y_1'(t) \\
y_2'(t) \\
y_3'(t)
\end{pmatrix} = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{pmatrix} \begin{pmatrix}
y_1(t) \\
y_2(t) \\
y_3(t)
\end{pmatrix}
\]

Denoting
\[
\begin{pmatrix}
y_1 \\
y_2 \\
y_3
\end{pmatrix} \quad \text{by} \quad Y
\]
and

\[
\begin{pmatrix}
y_1' \\
y_2' \\
y_3'
\end{pmatrix}
\text{ by } Y'
\]

the system can be written

\[Y' = AY\] (5.5.1)

where \( A \) is the \( 3 \times 3 \) matrix in the first equation. This system is called a \( 3 \times 3 \) system. An \( n \times n \) system can be written as Eq. (5.5.1) if we have

\[
Y = \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}
\]

and \( A \) is an \( n \times n \) matrix.

Such systems arise in simple diffusion problems in bioengineering. Consider two cells (Fig. 5.5.1) which contain equal volumes \( V \) of a salt solution. The boundary between the cells is a semipermeable membrane. Let \( y_1(t) \) and \( y_2(t) \) represent the amount of salt dissolved in cells 1 and cells 2, respectively. A reasonable assumption is that the rate of change of the amount of salt in a cell due to passage through the membrane is proportional to the difference in the concentrations in the cells. This leads to the equations

\[
\frac{y'_1(t)}{V} = k \left( \frac{y_2(t)}{V} - \frac{y_1(t)}{V} \right)
\]

Rate of change of concentration

Proportionality constant (\( > 0 \))

and the system

\[
\frac{y'_2(t)}{V} = k \left( \frac{y_1(t)}{V} - \frac{y_2(t)}{V} \right)
\]

and the system

\[
Y' = \begin{pmatrix}
-k & +k \\
+ & -k
\end{pmatrix} Y
\]
(we canceled the $V$’s). The matrix of the system is symmetric. This happens often in applications.

A first-order system such as Eq. (5.5.1), once derived as a model for a physical problem, must be solved. That is, functions $y_1(t), y_2(t), \ldots, y_n(t)$ must be found which, when substituted into (5.5.1), give a true mathematical equation.

**Example 1.** Show that

$$Y = \begin{pmatrix} ce^{-2t} \\ -ce^{-2t} \end{pmatrix}$$

is a solution of

$$Y' = \begin{pmatrix} -1 & +1 \\ +1 & -1 \end{pmatrix} Y$$

Show that

$$V = \begin{pmatrix} c \\ c \end{pmatrix}$$

is also a solution.

**Solution** We have

$$Y' = \begin{pmatrix} -2ce^{-2t} \\ +2ce^{-2t} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 & +1 \\ +1 & -1 \end{pmatrix} Y = \begin{pmatrix} -2ce^{-2t} \\ +2ce^{-2t} \end{pmatrix}$$

Therefore,

$$Y = \begin{pmatrix} ce^{-2t} \\ -ce^{-2t} \end{pmatrix}$$

is a solution.

For

$$V = \begin{pmatrix} c \\ c \end{pmatrix}$$

we obtain

$$V' = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} c \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Therefore,

$$V = \begin{pmatrix} c \\ c \end{pmatrix}$$

is a solution also.
The solutions given in Example 1 are closely related to the eigenvalues and eigenvectors of
\[
\begin{pmatrix}
-1 & 1 \\
1 & -1
\end{pmatrix}
\]

**Example 2.** Find the eigenvalues and eigenvectors of
\[
A = \begin{pmatrix}
-1 & 1 \\
1 & -1
\end{pmatrix}
\]
Relate them to the solutions of \(Y' = AY\) given in Example 1.

**Solution** The characteristic polynomial of \(A\) is \(p(\lambda) = \lambda^2 + 2\lambda\) which has roots \(\lambda_1 = -2\) and \(\lambda_2 = 0\). The eigenpairs are
\[
\left(-2, \begin{pmatrix} k \\ -k \end{pmatrix}\right) \quad \text{and} \quad \left(0, \begin{pmatrix} j \\ j \end{pmatrix}\right)
\]
The first eigenvalue is the exponent of the exponential function \(e^{-2t}\). If we multiply the eigenvector for \(-2\) by \(e^{-2t}\), we obtain the first solution in Example 1. Using the second eigenpair, if we multiply
\[
e^{0t} \begin{pmatrix} j \\ j \end{pmatrix} = \begin{pmatrix} j \\ j \end{pmatrix}
\]
we obtain the second solution in Example 1.

The results of Example 2 indicate that if \((\lambda, X)\) is an eigenpair of \(A\), then
\[
e^{\lambda t}X
\]
is a solution of \(Y' = AY\). This is basically true.

**Theorem 5.5.1.** Let \(A\) be a matrix with real entries. Let \(X\) be an eigenvector corresponding to the real eigenvalue \(\lambda\) of \(A\). The vector \(Y = e^{\lambda t}X\) is a solution of \(Y' = AY\) for all \(t\).

**Proof.** Let \(Y = e^{\lambda t}X\). Then \(Y' = \lambda e^{\lambda t}X\), since \(X\) is a constant vector. However, \(AY = A(e^{\lambda t}X) = e^{\lambda t}(AX) = e^{\lambda t}\lambda X = \lambda e^{\lambda t}X\). Thus \(Y' = AY\). \(\square\)

**Theorem 5.5.2.** If \(A\) is diagonalizable, then \(Y' = AY\) has \(n\) linearly independent solutions of the form in Theorem 5.5.1.
Theorem 5.5.3. If \( Y_1 \) and \( Y_2 \) are two solutions of \( Y' = AY \), then \( c_1 Y_1 + c_2 Y_2 \) is a solution also.

Proof. If \( Y' = AY \) and \( Y' = AY \), then

\[
A(c_1 Y_1 + c_2 Y_2) = c_1 AY_1 + c_2 AY_2 = c_1 Y'_1 + c_2 Y'_2 \\
= (c_1 Y_1 + c_2 Y_2)'
\]

Therefore, \( c_1 Y_1 + c_2 Y_2 \) is a solution.

These theorems tell us that if we solve the eigenproblem for a diagonalizable matrix \( A \), then we have solved

\[ Y' = AY \]

Example 3. Solve the diffusion problem from the beginning of this section:

\[ Y = \begin{pmatrix} -k & k \\ k & -k \end{pmatrix} Y \]

Solution. The characteristic polynomial of

\[ A = \begin{pmatrix} -k & k \\ k & -k \end{pmatrix} \]

is \( p(\lambda) = \lambda^2 + 2k\lambda \), so the eigenvalues are 0 and \(-2k\). Since

\[ \left( 0, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \quad \text{and} \quad \left( -2k, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) \]

are eigenpairs, we have

\[ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e^{-2kt} \\ -e^{-2kt} \end{pmatrix} \]

as independent solutions. By Theorem 5.5.3 a general solution of the system is

\[ c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} e^{-2kt} \\ -e^{-2kt} \end{pmatrix} = \begin{pmatrix} c_1 + c_2 e^{-2kt} \\ c_1 - c_2 e^{-2kt} \end{pmatrix} \]
5.5. POSTCALCULUS

Since the solution of the diffusion problem is (and it is the most general solution, a fact we will not prove)

\[
\begin{pmatrix}
    y_1(t) \\
    y_2(t)
\end{pmatrix}
= \begin{pmatrix}
    c_1 + c_2 e^{-2kt} \\
    c_1 - c_2 e^{-2kt}
\end{pmatrix}
\]

we note that as \( t \to +\infty \),

\[
\begin{pmatrix}
    y_1 \\
    y_2
\end{pmatrix}
\to \begin{pmatrix}
    c_1 \\
    c_1
\end{pmatrix}
\]

That is, as \( t \to \infty \), the diffusion system achieves the equilibrium state of equal concentrations of salt in each compartment. This makes sense and gives us confidence that the mathematical model of our differential equation may be an accurate model for the diffusion system.

Diffusion problems ordinarily include an initial condition: the amount of salt in each compartment at the starting time \( (t = 0) \). For example, if we began the system with \( y_1(0) = 2 \) and \( y_2(0) = 1 \), the solution would have to satisfy

\[
\begin{pmatrix}
    y_1(0) \\
    y_2(0)
\end{pmatrix}
= \begin{pmatrix}
    2 \\
    1
\end{pmatrix}
\]

This means, by putting \( t = 0 \) into Eq. (5.5.2), that

\[
\begin{pmatrix}
    c_1 + c_2 \\
    c_1 - c_2
\end{pmatrix}
= \begin{pmatrix}
    2 \\
    1
\end{pmatrix}
\]

or \( c_1 = \frac{3}{2} \) and \( c_2 = \frac{1}{2} \). Therefore, the solution of the diffusion problem would be

\[
\begin{pmatrix}
    y_1(t) \\
    y_2(t)
\end{pmatrix}
= \begin{pmatrix}
    \frac{3}{2} + \frac{1}{2} e^{-2kt} \\
    \frac{3}{2} - \frac{1}{2} e^{-2kt}
\end{pmatrix}
\]

This solution reflects some physical principles:

1. As \( t \to \infty \), the salt will balance between the two compartments. As \( t \to \infty \),

\[
\begin{pmatrix}
    y_1(t) \\
    y_2(t)
\end{pmatrix}
\to \begin{pmatrix}
    \frac{3}{2} \\
    \frac{1}{2}
\end{pmatrix}
\]

Since the volume of each cell is the same, the concentrations are equal.
2. The total amount of salt is conserved. At the beginning we have $2+1 = 3$ units of salt. For all $t$,

$$y_1(t) + y_2(t) = \frac{3}{2} + \frac{1}{2}e^{-2kt} + \frac{3}{2} - \frac{1}{2}e^{-2kt} = 3 \text{ units of salt}$$

3. The first compartment will always have more salt than the second. In our solution $y_1(t) > y_2(t)$ for any finite time $t$ since $\frac{1}{2}e^{-2kt} > -\frac{1}{2}e^{-2kt}$.

4. The amount of salt in compartment 1 (the saltier one at the beginning) will always be decreasing. Since $y_1(t) = \frac{3}{2} + \frac{1}{2}e^{-2kt}$, we have the derivative

$$y_1'(t) = -ke^{-2kt} < 0 \quad \text{for all } t$$

Thus $y_1(t)$ is decreasing.

5. The amount of salt in compartment 2 will always be increasing. This is shown in the same way as principle 4 was shown.

6. As the permeability increases ($k$ gets larger), the balancing of the salt occurs more quickly. If $k_1 > k_2$, then

$$e^{-2k_1t} < e^{-2k_2t} \quad \text{and} \quad \frac{3}{2} + \frac{1}{2}e^{-2k_1t} < \frac{3}{2} + \frac{1}{2}e^{-2k_2t}$$

(See Fig. 5.5.2.)

Since the solution to the system of differential equations adheres to these physical intuitions we have about the system, we have confidence in the mathematical model. Of course, conducting an experiment and making observations are the ultimate test for the model. It turns out that the model is good for dilute solutions; moreover, solutes other than salt can be modeled.

For large systems, direct numerical methods are often used instead of attempting diagonalization of $A$. Diagonalization of $A$ has importance in finding canonical coordinates for a system of differential equations.

To summarize: The general solution to

$$Y' = AY$$

where $A$ is diagonalizable with distinct eigenvalues, can be written as

$$Y = c_1e^{\lambda_1t}X_1 + c_2e^{\lambda_2t}X_2 + \cdots + c_ne^{\lambda_nt}X_n$$
where \(c_1, \ldots, c_n\) are arbitrary constants, \(\lambda_1, \ldots, \lambda_n\) are the eigenvalues of \(A\), and \(X_1, \ldots, X_n\) are the independent eigenvectors of \(A\). The constants \(c_1, \ldots, c_n\) are determined exactly when an initial condition

\[
Y(0) = \begin{pmatrix} y_1(0) \\ \vdots \\ y_n(0) \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}
\]

gives.

Finally, if we allow complex solutions of \(\det(A - \lambda I) = 0\) to be used for eigenvalues, we can find solutions which are sine and cosine functions. For example,

\[
\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}
\]

has solution

\[
Y = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}
\]

Note that the eigenvalues of

\[
\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}
\]

are \(i\) and \(-i\). The connection to the solution is that

\[e^{it} = \cos t + i \sin t\]

a fundamental identity for the analysis of differential equations—especially as models of vibrating systems in mechanics.

**Vibrations** The system of trolleys in Fig. 5.5.3 will vibrate if they are pulled and released. However, only certain vibrations can be sustained without outside forces. The frequencies of these vibrations are called the fundamental frequencies of the system. These frequencies can be found by solving the eigenvalue problem for a certain matrix. The matrix is obtained as follows: With \(y_1(t)\) and \(y_2(t)\) as the displacements of trolley 1 and trolley 2, respectively, from equilibrium (see Fig. 5.5.4), the equations of motion (with no external forces) are

\[
y''_1 = -2k \frac{k}{m} y_1 + \frac{k}{m} y_2
\]

\[
y''_2 = \frac{k}{m} y_1 - 2 \frac{k}{m} y_2
\]
Assuming solutions of the form $y_1 = A \cos \omega t + B \sin \omega t$ and $y_2 = C \cos \omega t + D \sin \omega t$, where $\omega$ is the frequency, and substituting we find

$$-\omega^2 y_1 = -2 \frac{k}{m} y_1 + \frac{k}{m} y_2$$
$$-\omega^2 y_2 = \frac{k}{m} y_1 - 2 \frac{k}{m} y_2$$

or

$$\begin{pmatrix} -2 \frac{k}{m} & \frac{k}{m} \\ \frac{k}{m} & -2 \frac{k}{m} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = -\omega^2 \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (5.5.3)$$

This last equation means that solutions of the form specified can exist only if the matrix (call it $A$) in Eq. (5.5.3) has negative eigenvalues; for if $\lambda$ is an eigenvalue, $\lambda = -\omega^2$ and $\omega$ is real.

The matrix $A$ has eigenpairs

$$\begin{pmatrix} -3 \frac{k}{m} & \frac{k}{m} \\ \frac{k}{m} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{k}{m} \\ \frac{k}{m} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The interpretations is, since $-\omega^2$ is $-3k/m$ or $-k/m$, that the characteristic frequencies are $\sqrt{3k/m}$ and $\sqrt{k/m}$. Also the opposite signs of the components in the first eigenvector indicate that $y_1 = -y_2$ (trolleys moving in opposite directions), and the same signs of the components in the second eigenvector indicate that $y_1 = y_2$ (trolleys moving in the same direction).

**Stability** For vibrating systems as well as diffusion systems in equilibrium, stability of equilibrium is an important consideration. Roughly speaking, an equilibrium state for a system described by systems of differential equations is a state for which the velocity of the components is zero. For example, a pendulum has two equilibrium states, as shown in Fig. 5.5.5. However, one equilibrium state is fundamentally different from the other. The second is **stable**: After a slight push the pendulum will return to its original position. In the first position a slight push will cause the pendulum to leave that equilibrium state, never to return.

Stability enters into the design of structures, chemical reactors, and machines. Often, several equilibrium configurations are theoretically possible in chemical reactors, but only the stable equilibrium can be maintained in practice.
5.5. *POSTCALCULUS*

Roughly speaking, for the types of systems we have briefly discussed, stability is guaranteed if the eigenvalues of matrix $A$ of the differential equation $Y'' = AY$ have negative real parts. This means that in both of our examples the system was stable. Rigorous discussions of stability can be found in many texts on ordinary differential equations.

**PROBLEMS 5.5**

1–7 Find the general solution to the system $Y' = AY$ for the matrices $A$ given in Probs. 1 to 7 of the last section, provided the given matrix has distinct eigenvalues.

Use the general solutions in Probs. 1 to 7 to solve the differential equations in Probs. 8 to 10 with initial conditions as given:

8. For $3 \times 3$ matrices $Y(0) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

9. For $2 \times 2$ matrices $Y(0) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

10. For $4 \times 4$ matrices $Y(0) = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$

11. Consider a three-cell diffusion system with cells of equal volume separated by semipermeable membranes, as shown. Suppose the diffusion coefficients are all equal to $k(>0)$. Show that the first-order system modeling the diffusion is

\[
\begin{align*}
y_1'(t) &= k[y_2(t) - y_1(t)] + k[y_3(t) - y_1(t)] \\
y_2'(t) &= k[y_1(t) - y_2(t)] + k[y_3(t) - y_2(t)] \\
y_3'(t) &= k[y(t) - y_3(t)] + k[y_2(t) - y_3(t)]
\end{align*}
\]

where $y_1$, $y_2$, and $y_3$ are the amounts of salt in compartments 1, 2, and 3, respectively. Show that the matrix of the system has an eigenvalue of multiplicity 2.
12. The system below has equations of motion (gravity neglected).

\[
\begin{pmatrix}
 y_1'' \\
 y_2''
\end{pmatrix} = \begin{pmatrix}
 -(k_1 + k_2) & k_2 \\
 m_1 & m_2 \\
 k_2 & -k_2 \\
 m_2 & m_2
\end{pmatrix} \begin{pmatrix}
 y_1 \\
 y_2
\end{pmatrix}
\]

where \( y_1(t) \) and \( y_2(t) \) are the displacements of masses \( m_1 \) and \( m_2 \), respectively, from equilibrium.

(a) For \( k_1 = k_2 = k \) and \( m_1 = m_2 = m \), find the characteristic frequencies.

(b) For \( k_1 = k_2 = k \) and \( m_1 = m_2 = m \), show that the eigenvalues of \( A \) are negative to show that the system is stable.

(c) If \( k_1 \neq k_2 \) and \( m_1 \neq m_2 \), is the system stable?

SUMMARY

Representations of a linear transformation by a diagonal matrix is advantageous but not possible for all transformations. The problem reduces to asking when a matrix can be diagonalized. An \( n \times n \) matrix can be diagonalized if and only if it possesses a set of \( n \) linearly independent eigenvectors. Thus the solution of the diagonalization problem depends on finding the eigenvalues and eigenvectors of the matrix. In applications, the eigenvalues are related to frequencies of vibration and stability of mechanical systems.

Real symmetric matrices are always orthogonally diagonalizable, and hermitian matrices are always unitarily diagonalizable. To carry out these diagonalizations may require the Gram-Schmidt procedure to construct an orthonormal basis of eigenvectors of the matrix.

As a result of the efforts in the first five chapters, we have presented solutions of the five basic problems of linear algebra:

1. Solution of linear equations. Gaussian elimination was the primary method of solution, and attention was given to numerical considerations.

2. The basis problem. Using the concepts of linear independence and dependence, we showed how to solve this problem in a constructive, "basis-building" way.
3. **The matrix representation problem.** Linear transformations can be represented by matrices; the solution involved the representation of images under the transformation of basis elements.

4. **The eigenvalue-eigenvector problem.** This problem was solved by using the methods for solving linear equations and determinants. The solution to this problem is critical to success in solving the last basic problem.

5. **The diagonalization problem.** Using the techniques of the solutions of the first four basic problems, we solved this problem which has many applications in science and engineering.

Certainly, a good knowledge of the solution methods for these five problems equips a person to go on to more advanced study of linear algebra. Meanwhile it gives one the tools to handle basic linear algebraic problems and methods in science, engineering, and numerical analysis. As the reader has noticed, the problems and examples of this chapter were constructed so that eigenvalues and eigenvectors were reasonably calculable by hand. In general, this is not the case so numerical methods have been devised to handle these problems. Two of these methods are discussed in the upcoming chapter.

### ADDITIONAL PROBLEMS

1. Is the set of all diagonalizable $n \times n$ matrices a subspace of $\mathcal{M}_{nn}$?

2. Let $A$ be an $n \times n$ diagonalizable matrix. Discuss the solvability of $AX_{n \times 1} = B$.

3. Find the characteristic polynomials of

$$
\begin{pmatrix}
0 & 1 \\
-a_1 & -a_2
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
-a_1 & -a_2 & -a_3
\end{pmatrix}
$$
CHAPTER 5. EIGENVALUES AND DIAGONALIZATION

4. Find the characteristic polynomial of

\[
\begin{pmatrix}
0 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 \\
-a_1 & -a_2 & -a_3 & -a_4 & \cdots & -a_n
\end{pmatrix}
\]

5. Let \( A_{n\times n} \) be defined by \( a_{ij} = 1 \) for all \( i \) and \( j \). Show that the characteristic equation is \((-\lambda)^{n-1}(n - \lambda) = 0\).

6. Suppose \( A \) and \( B \) are \( n \times n \) matrices and \( A = PD\lambda P^{-1} \) and \( B = PE\lambda P^{-1} \), where \( D \) and \( E \) are diagonal matrices. Compare \( AB \) and \( BA \).

7. If a diagonalizable matrix is nilpotent what can you say about its trace?

8. Let

\[
A = \begin{pmatrix}
-(b + a) & a & b \\
a & -(a + c) & c \\
b & c & -(b + c)
\end{pmatrix}
\]

Show that \( A \) has an eigenvalue of zero, with corresponding eigenvector

\[
\begin{pmatrix}
1 \\
1 \\
1
\end{pmatrix}
\]

What can you say about the other two eigenvalue?

9. A vector \( X_{n\times1} \) is called a root vector of order \( k \) for the matrix \( A_{n\times n} \) if there exists a number \( \lambda \) such that

\[
(A - \lambda I)^{k-1}X \neq 0 \quad \text{and} \quad (A - \lambda I)^kX = 0
\]

Root vectors are useful in studying systems of differential equations. Show that

\[
\begin{pmatrix}
2 \\
0 \\
1
\end{pmatrix}
\]
is a root vector of order 3 for

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

with \( \lambda = 0 \).

10. Show that

\[
\begin{pmatrix}
3 & 1 \\
2 & 1
\end{pmatrix}
\]

has no root vector of order 2.

11. If \( A \) is real and diagonalizable with all positive eigenvalues, then does \( A \) have a square root? That is, does there exist a real matrix \( B \) with \( A = B^2 \)?

12. In mathematical physics, lagrangian equations of motion are sometimes used to study vibrating systems. These equations generally involve two quadratic forms, and so we have the problem of diagonalizing the matrices of both quadratic forms at the same time. If \( A \) and \( B \) are real symmetric matrices and one of \( A \) and \( B \) is positive definite, then the matrices can be simultaneously diagonalized. That is, there exists a matrix \( P \) such that \( P^T AP \) and \( P^T BP \) are diagonal. Show by computation of \( P^T AP \) and \( P^T BP \) that

\[
P = \begin{pmatrix}
1 & 1 \\
2\sqrt{2} & 2\sqrt{2} \\
1 & 1 \\
2 & 2
\end{pmatrix}
\]

simultaneously diagonalizes

\[
\begin{pmatrix}
a & a \\
2 & 2
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
b & 0 \\
0 & b
\end{pmatrix}
\]

13. Consider the vibrating system with damper (like a shock absorber) in Fig. AP5.13. The system of differential equations for this system has matrix

\[
\begin{pmatrix}
0 & 1 \\
-c & -k \\
m & m
\end{pmatrix}
\]
CHAPTER 5. EIGENVALUES AND DIAGONALIZATION

What conditions on \( k, c, \) and \( m \) guarantee complex eigenvalues? This corresponds to the case of underdamping or light damping—the mass oscillates, with amplitude going to zero as time goes on. The imaginary part of the eigenvalue gives the frequency of oscillation.

14. In Prob. 12 of the Additional Problems in Chap. 4 we presented a matrix associated with age distributions in a population. An eigenvector associated with the largest positive eigenvalue (if such an eigenvalue exists) is called a **stable age distribution**. The components of the eigenvector give the relative proportions of the age groups. For example, the matrix

\[
A = \begin{pmatrix}
0 & 0 & 6 \\
\frac{1}{2} & 0 & 0 \\
0 & \frac{1}{2} & 0
\end{pmatrix}
\]

represents a species with life span of 3 years (hence the \( 3 \times 3 \) matrix) which produces six offspring in the third year (the entry \( a_{13} \)) but produces none when younger. Show that \( A \) has eigenvector

\[
\begin{pmatrix}
6 \\
3 \\
1
\end{pmatrix}
\]

corresponding to its largest positive eigenvalue. This means that in the stable age distribution, the ratio of the age groups 0 to 1, 1 to 2, and 2 to 3 years in 6:3:1.

15. The circuit in Fig. AP5.15 leads to consideration of a system of differential equations involving the matrix. Under what conditions on \( L, R, \) and \( C \) does \( A \) have complex eigenvalues with nonzero imaginary part? In that case, show that the real parts of the eigenvalues are negative. This means that the currents involved all die to zero in an oscillatory fashion.