8.2 Orthogonal Diagonalization

Recall (Theorem 5.5.3) that an $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. Moreover, the matrix P with these eigenvectors as columns is a diagonalizing matrix for A, that is

$$P^{-1}AP \text{ is diagonal.}$$

As we have seen, the really nice bases of $\mathbb{R}^n$ are the orthogonal ones, so a natural question is: which $n \times n$ matrices have an *orthogonal* basis of eigenvectors? These turn out to be precisely the symmetric matrices, and this is the main result of this section.

Before proceeding, recall that an orthogonal set of vectors is called *orthonormal* if $\|\mathbf{v}\| = 1$ for each vector $\mathbf{v}$ in the set, and that any orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ can be "*normalized*", that is converted into an orthonormal set $\{\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \dots, \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k\}$. In particular, if a matrix A has n orthogonal eigenvectors, they can (by normalizing) be taken to be orthonormal. The corresponding diagonalizing matrix P has orthonormal columns, and such matrices are very easy to invert.

Theorem 8.2.1

The following conditions are equivalent for an $n \times n$ matrix P.

1. *P is invertible and $P^{-1} = P^T$.*
2. *The rows of P are orthonormal.*
3. *The columns of P are orthonormal.*

Proof. First recall that condition (1) is equivalent to $PP^T = I$ by Corollary 2.4.1 of Theorem 2.4.5. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ denote the rows of P. Then $\mathbf{x}_j^T$ is the jth column of P^T, so the (i, j)-entry of PP^T is $\mathbf{x}_i \cdot \mathbf{x}_j$. Thus $PP^T = I$ means that $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ if $i \neq j$ and $\mathbf{x}_i \cdot \mathbf{x}_j = 1$ if $i = j$. Hence condition (1) is equivalent to (2). The proof of the equivalence of (1) and (3) is similar. □

Definition 8.3 Orthogonal Matrices

An $n \times n$ matrix P is called an ***orthogonal matrix***[2] *if it satisfies one (and hence all) of the conditions in Theorem 8.2.1.*

Example 8.2.1

The rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal for any angle θ.

These orthogonal matrices have the virtue that they are easy to invert—simply take the transpose. But they have many other important properties as well. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, we will prove

[2] In view of (2) and (3) of Theorem 8.2.1, *orthonormal matrix* might be a better name. But *orthogonal matrix* is standard.

(Theorem 10.4.3) that T is distance preserving if and only if its matrix is orthogonal. In particular, the matrices of rotations and reflections about the origin in $\mathbb{R}^2$ and $\mathbb{R}^3$ are all orthogonal (see Example 8.2.1).

It is not enough that the rows of a matrix A are merely orthogonal for A to be an orthogonal matrix. Here is an example.

Example 8.2.2

The matrix $\left[\begin{array}{rrr} 2 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{array}\right]$ has orthogonal rows but the columns are not orthogonal. However, if the rows are normalized, the resulting matrix $\left[\begin{array}{ccc} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{array}\right]$ is orthogonal (so the columns are now orthonormal as the reader can verify).

Example 8.2.3

If P and Q are orthogonal matrices, then PQ is also orthogonal, as is $P^{-1} = P^T$.

Solution. P and Q are invertible, so PQ is also invertible and

$$(PQ)^{-1} = Q^{-1}P^{-1} = Q^T P^T = (PQ)^T$$

Hence PQ is orthogonal. Similarly,

$$(P^{-1})^{-1} = P = (P^T)^T = (P^{-1})^T$$

shows that P^{-1} is orthogonal.

Definition 8.4 Orthogonally Diagonalizable Matrices

An $n \times n$ matrix A is said to be ***orthogonally diagonalizable*** *when an orthogonal matrix P can be found such that $P^{-1}AP = P^T AP$ is diagonal.*

This condition turns out to characterize the symmetric matrices.

Theorem 8.2.2: Principal Axes Theorem

The following conditions are equivalent for an $n \times n$ matrix A.

1. *A has an orthonormal set of n eigenvectors.*
2. *A is orthogonally diagonalizable.*
3. *A is symmetric.*

Proof. (1) $\Leftrightarrow$ (2). Given (1), let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be orthonormal eigenvectors of A. Then $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$ is orthogonal, and $P^{-1}AP$ is diagonal by Theorem 3.3.4. This proves (2). Conversely, given (2) let $P^{-1}AP$ be diagonal where P is orthogonal. If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are the columns of P then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ that consists of eigenvectors of A by Theorem 3.3.4. This proves (1).

(2) $\Rightarrow$ (3). If $P^TAP = D$ is diagonal, where $P^{-1} = P^T$, then $A = PDP^T$. But $D^T = D$, so this gives $A^T = P^{TT}D^TP^T = PDP^T = A$.

(3) $\Rightarrow$ (2). If A is an $n \times n$ symmetric matrix, we proceed by induction on n. If $n = 1$, A is already diagonal. If $n > 1$, assume that (3) $\Rightarrow$ (2) for $(n-1) \times (n-1)$ symmetric matrices. By Theorem 5.5.7 let λ_1 be a (real) eigenvalue of A, and let $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$, where $\|\mathbf{x}_1\| = 1$. Use the Gram-Schmidt algorithm to find an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ for $\mathbb{R}^n$. Let $P_1 = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$, so P_1 is an orthogonal matrix and $P_1^TAP_1 = \begin{bmatrix} \lambda_1 & B \\ 0 & A_1 \end{bmatrix}$ in block form by Lemma 5.5.2. But $P_1^TAP_1$ is symmetric (A is), so it follows that $B = 0$ and A_1 is symmetric. Then, by induction, there exists an $(n-1) \times (n-1)$ orthogonal matrix Q such that $Q^TA_1Q = D_1$ is diagonal. Observe that $P_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal, and compute:

$$\begin{aligned}
(P_1P_2)^TA(P_1P_2) &= P_2^T(P_1^TAP_1)P_2 \\
&= \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & 0 \\ 0 & D_1 \end{bmatrix}
\end{aligned}$$

is diagonal. Because P_1P_2 is orthogonal, this proves (2). □

A set of orthonormal eigenvectors of a symmetric matrix A is called a set of **principal axes** for A. The name comes from geometry, and this is discussed in Section 8.9. Because the eigenvalues of a (real) symmetric matrix are real, Theorem 8.2.2 is also called the **real spectral theorem**, and the set of distinct eigenvalues is called the **spectrum** of the matrix. In full generality, the spectral theorem is a similar result for matrices with complex entries (Theorem 8.7.8).

Example 8.2.4

Find an orthogonal matrix P such that $P^{-1}AP$ is diagonal, where $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$.

Solution. The characteristic polynomial of A is (adding twice row 1 to row 2):

$$c_A(x) = \det \begin{bmatrix} x-1 & 0 & 1 \\ 0 & x-1 & -2 \\ 1 & -2 & x-5 \end{bmatrix} = x(x-1)(x-6)$$

Thus the eigenvalues are $\lambda = 0$, 1, and 6, and corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

respectively. Moreover, by what appears to be remarkably good luck, these eigenvectors are *orthogonal*. We have $\|\mathbf{x}_1\|^2 = 6$, $\|\mathbf{x}_2\|^2 = 5$, and $\|\mathbf{x}_3\|^2 = 30$, so

$$P = \left[\begin{array}{ccc} \frac{1}{\sqrt{6}}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{\sqrt{30}}\mathbf{x}_3 \end{array} \right] = \frac{1}{\sqrt{30}} \left[\begin{array}{ccc} \sqrt{5} & 2\sqrt{6} & -1 \\ -2\sqrt{5} & \sqrt{6} & 2 \\ \sqrt{5} & 0 & 5 \end{array} \right]$$

is an orthogonal matrix. Thus $P^{-1} = P^T$ and

$$P^T AP = \left[\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{array} \right]$$

by the diagonalization algorithm.

Actually, the fact that the eigenvectors in Example 8.2.4 are orthogonal is no coincidence. Theorem 5.5.4 guarantees they are linearly independent (they correspond to distinct eigenvalues); the fact that the matrix is *symmetric* implies that they are orthogonal. To prove this we need the following useful fact about symmetric matrices.

Theorem 8.2.3

If A is an $n \times n$ symmetric matrix, then

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$$

for all columns $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.[3]

Proof. Recall that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$ for all columns $\mathbf{x}$ and $\mathbf{y}$. Because $A^T = A$, we get

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

□

Theorem 8.2.4

If A is a symmetric matrix, then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof. Let $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \mu\mathbf{y}$, where $\lambda \neq \mu$. Using Theorem 8.2.3, we compute

$$\lambda(\mathbf{x} \cdot \mathbf{y}) = (\lambda\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y}) = \mathbf{x} \cdot (\mu\mathbf{y}) = \mu(\mathbf{x} \cdot \mathbf{y})$$

Hence $(\lambda - \mu)(\mathbf{x} \cdot \mathbf{y}) = 0$, and so $\mathbf{x} \cdot \mathbf{y} = 0$ because $\lambda \neq \mu$. □

[3] The converse also holds (Exercise 8.2.15).

Now the procedure for diagonalizing a symmetric $n \times n$ matrix is clear. Find the distinct eigenvalues (all real by Theorem 5.5.7) and find orthonormal bases for each eigenspace (the Gram-Schmidt algorithm may be needed). Then the set of all these basis vectors is orthonormal (by Theorem 8.2.4) and contains n vectors. Here is an example.

Example 8.2.5

Orthogonally diagonalize the symmetric matrix $A = \left[\begin{array}{rrr} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{array}\right]$.

Solution. The characteristic polynomial is

$$c_A(x) = \det \left[\begin{array}{ccc} x-8 & 2 & -2 \\ 2 & x-5 & -4 \\ -2 & -4 & x-5 \end{array}\right] = x(x-9)^2$$

Hence the distinct eigenvalues are 0 and 9 of multiplicities 1 and 2, respectively, so $\dim(E_0) = 1$ and $\dim(E_9) = 2$ by Theorem 5.5.6 (A is diagonalizable, being symmetric). Gaussian elimination gives

$$E_0(A) = \text{span}\,\{\mathbf{x}_1\}, \quad \mathbf{x}_1 = \left[\begin{array}{r} 1 \\ 2 \\ -2 \end{array}\right], \qquad \text{and} \qquad E_9(A) = \text{span}\left\{ \left[\begin{array}{r} -2 \\ 1 \\ 0 \end{array}\right], \left[\begin{array}{r} 2 \\ 0 \\ 1 \end{array}\right] \right\}$$

The eigenvectors in E_9 are both orthogonal to $\mathbf{x}_1$ as Theorem 8.2.4 guarantees, but not to each other. However, the Gram-Schmidt process yields an orthogonal basis

$$\{\mathbf{x}_2, \mathbf{x}_3\} \text{ of } E_9(A) \qquad \text{where} \qquad \mathbf{x}_2 = \left[\begin{array}{r} -2 \\ 1 \\ 0 \end{array}\right] \text{ and } \mathbf{x}_3 = \left[\begin{array}{r} 2 \\ 4 \\ 5 \end{array}\right]$$

Normalizing gives orthonormal vectors $\{\frac{1}{3}\mathbf{x}_1, \frac{1}{\sqrt{5}}\mathbf{x}_2, \frac{1}{3\sqrt{5}}\mathbf{x}_3\}$, so

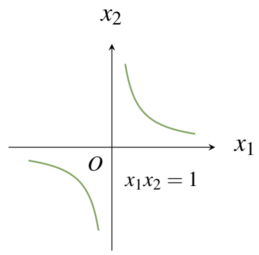
$$P = \left[\begin{array}{ccc} \frac{1}{3}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{3\sqrt{5}}\mathbf{x}_3 \end{array}\right] = \frac{1}{3\sqrt{5}} \left[\begin{array}{rrr} \sqrt{5} & -6 & 2 \\ 2\sqrt{5} & 3 & 4 \\ -2\sqrt{5} & 0 & 5 \end{array}\right]$$

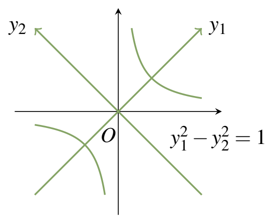
is an orthogonal matrix such that $P^{-1}AP$ is diagonal.

It is worth noting that other, more convenient, diagonalizing matrices P exist. For example, $\mathbf{y}_2 = \left[\begin{array}{r} 2 \\ 1 \\ 2 \end{array}\right]$ and $\mathbf{y}_3 = \left[\begin{array}{r} -2 \\ 2 \\ 1 \end{array}\right]$ lie in $E_9(A)$ and they are orthogonal. Moreover, they both have norm 3 (as does $\mathbf{x}_1$), so

$$Q = \left[\begin{array}{ccc} \frac{1}{3}\mathbf{x}_1 & \frac{1}{3}\mathbf{y}_2 & \frac{1}{3}\mathbf{y}_3 \end{array}\right] = \frac{1}{3} \left[\begin{array}{rrr} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{array}\right]$$

is a nicer orthogonal matrix with the property that $Q^{-1}AQ$ is diagonal.





If A is symmetric and a set of orthogonal eigenvectors of A is given, the eigenvectors are called principal axes of A. The name comes from geometry. An expression $q = ax_1^2 + bx_1x_2 + cx_2^2$ is called a **quadratic form** in the variables x_1 and x_2, and the graph of the equation $q = 1$ is called a **conic** in these variables. For example, if $q = x_1x_2$, the graph of $q = 1$ is given in the first diagram.

But if we introduce new variables y_1 and y_2 by setting $x_1 = y_1 + y_2$ and $x_2 = y_1 - y_2$, then q becomes $q = y_1^2 - y_2^2$, a diagonal form with no cross term y_1y_2 (see the second diagram). Because of this, the y_1 and y_2 axes are called the principal axes for the conic (hence the name). Orthogonal diagonalization provides a systematic method for finding principal axes. Here is an illustration.

Example 8.2.6

Find principal axes for the quadratic form $q = x_1^2 - 4x_1x_2 + x_2^2$.

Solution. In order to utilize diagonalization, we first express q in matrix form. Observe that

$$q = \left[\begin{array}{cc} x_1 & x_2 \end{array}\right] \left[\begin{array}{rr} 1 & -4 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right]$$

The matrix here is not symmetric, but we can remedy that by writing

$$q = x_1^2 - 2x_1x_2 - 2x_2x_1 + x_2^2$$

Then we have

$$q = \left[\begin{array}{cc} x_1 & x_2 \end{array}\right] \left[\begin{array}{rr} 1 & -2 \\ -2 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \mathbf{x}^T A \mathbf{x}$$

where $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right]$ and $A = \left[\begin{array}{rr} 1 & -2 \\ -2 & 1 \end{array}\right]$ is symmetric. The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$, with corresponding (orthogonal) eigenvectors $\mathbf{x}_1 = \left[\begin{array}{r} 1 \\ -1 \end{array}\right]$ and $\mathbf{x}_2 = \left[\begin{array}{c} 1 \\ 1 \end{array}\right]$. Since $\|\mathbf{x}_1\| = \|\mathbf{x}_2\| = \sqrt{2}$, so

$$P = \frac{1}{\sqrt{2}} \left[\begin{array}{rr} 1 & 1 \\ -1 & 1 \end{array}\right] \text{ is orthogonal and } P^T AP = D = \left[\begin{array}{rr} 3 & 0 \\ 0 & -1 \end{array}\right]$$

Now define new variables $\left[\begin{array}{c} y_1 \\ y_2 \end{array}\right] = \mathbf{y}$ by $\mathbf{y} = P^T \mathbf{x}$, equivalently $\mathbf{x} = P\mathbf{y}$ (since $P^{-1} = P^T$). Hence

$$y_1 = \frac{1}{\sqrt{2}}(x_1 - x_2) \quad \text{and} \quad y_2 = \frac{1}{\sqrt{2}}(x_1 + x_2)$$

In terms of y_1 and y_2, q takes the form

$$q = \mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T (P^T AP)\mathbf{y} = \mathbf{y}^T D\mathbf{y} = 3y_1^2 - y_2^2$$

Note that $\mathbf{y} = P^T \mathbf{x}$ is obtained from $\mathbf{x}$ by a counterclockwise rotation of $\frac{\pi}{4}$ (see Theorem 2.4.6).

Observe that the quadratic form q in Example 8.2.6 can be diagonalized in other ways. For example

$$q = x_1^2 - 4x_1x_2 + x_2^2 = z_1^2 - \tfrac{1}{3}z_2^2$$

where $z_1 = x_1 - 2x_2$ and $z_2 = 3x_2$. We examine this more carefully in Section 8.9.

If we are willing to replace "diagonal" by "upper triangular" in the principal axes theorem, we can weaken the requirement that A is symmetric to insisting only that A has real eigenvalues.

Theorem 8.2.5: Triangulation Theorem

If A is an $n \times n$ matrix with n real eigenvalues, an orthogonal matrix P exists such that $P^T AP$ is upper triangular.[4]

Proof. We modify the proof of Theorem 8.2.2. If $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ where $\|\mathbf{x}_1\| = 1$, let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ be an orthonormal basis of $\mathbb{R}^n$, and let $P_1 = \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array}\right]$. Then P_1 is orthogonal and $P_1^T AP_1 = \left[\begin{array}{cc} \lambda_1 & B \\ 0 & A_1 \end{array}\right]$ in block form. By induction, let $Q^T A_1 Q = T_1$ be upper triangular where Q is of size $(n-1) \times (n-1)$ and orthogonal. Then $P_2 = \left[\begin{array}{cc} 1 & 0 \\ 0 & Q \end{array}\right]$ is orthogonal, so $P = P_1P_2$ is also orthogonal and $P^T AP = \left[\begin{array}{cc} \lambda_1 & BQ \\ 0 & T_1 \end{array}\right]$ is upper triangular. □

The proof of Theorem 8.2.5 gives no way to construct the matrix P. However, an algorithm will be given in Section 11.1 where an improved version of Theorem 8.2.5 is presented. In a different direction, a version of Theorem 8.2.5 holds for an arbitrary matrix with complex entries (Schur's theorem in Section 8.7).

As for a diagonal matrix, the eigenvalues of an upper triangular matrix are displayed along the main diagonal. Because A and $P^T AP$ have the same determinant and trace whenever P is orthogonal, Theorem 8.2.5 gives:

Corollary 8.2.1

If A is an $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ (possibly not all distinct), then $\det A = \lambda_1\lambda_2 \dots \lambda_n$ *and* $\operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

This corollary remains true even if the eigenvalues are not real (using Schur's theorem).

Exercises for 8.2

Exercise 8.2.1 Normalize the rows to make each of the following matrices orthogonal.

a. $A = \left[\begin{array}{rr} 1 & 1 \\ -1 & 1 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 3 & -4 \\ 4 & 3 \end{array}\right]$

c. $A = \left[\begin{array}{rr} 1 & 2 \\ -4 & 2 \end{array}\right]$

d. $A = \left[\begin{array}{rr} a & b \\ -b & a \end{array}\right]$, $(a, b) \neq (0, 0)$

[4] There is also a lower triangular version.

e. $A = \left[\begin{array}{ccc} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 2 \end{array}\right]$

f. $A = \left[\begin{array}{rrr} 2 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{array}\right]$

g. $A = \left[\begin{array}{rrr} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{array}\right]$

h. $A = \left[\begin{array}{rrr} 2 & 6 & -3 \\ 3 & 2 & 6 \\ -6 & 3 & 2 \end{array}\right]$

Exercise 8.2.2 If P is a triangular orthogonal matrix, show that P is diagonal and that all diagonal entries are 1 or -1.

Exercise 8.2.3 If P is orthogonal, show that kP is orthogonal if and only if $k = 1$ or $k = -1$.

Exercise 8.2.4 If the first two rows of an orthogonal matrix are $(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and $(\frac{2}{3}, \frac{1}{3}, \frac{-2}{3})$, find all possible third rows.

Exercise 8.2.5 For each matrix A, find an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

a. $A = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 1 & -1 \\ -1 & 1 \end{array}\right]$

c. $A = \left[\begin{array}{ccc} 3 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 5 \end{array}\right]$

d. $A = \left[\begin{array}{ccc} 3 & 0 & 7 \\ 0 & 5 & 0 \\ 7 & 0 & 3 \end{array}\right]$

e. $A = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right]$

f. $A = \left[\begin{array}{rrr} 5 & -2 & -4 \\ -2 & 8 & -2 \\ -4 & -2 & 5 \end{array}\right]$

g. $A = \left[\begin{array}{cccc} 5 & 3 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 1 & 7 \end{array}\right]$

h. $A = \left[\begin{array}{rrrr} 3 & 5 & -1 & 1 \\ 5 & 3 & 1 & -1 \\ -1 & 1 & 3 & 5 \\ 1 & -1 & 5 & 3 \end{array}\right]$

Exercise 8.2.6 Consider $A = \left[\begin{array}{ccc} 0 & a & 0 \\ a & 0 & c \\ 0 & c & 0 \end{array}\right]$ where one of $a,\ c \neq 0$. Show that $c_A(x) = x(x-k)(x+k)$, where $k = \sqrt{a^2+c^2}$ and find an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Exercise 8.2.7 Consider $A = \left[\begin{array}{ccc} 0 & 0 & a \\ 0 & b & 0 \\ a & 0 & 0 \end{array}\right]$. Show that $c_A(x) = (x-b)(x-a)(x+a)$ and find an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Exercise 8.2.8 Given $A = \left[\begin{array}{cc} b & a \\ a & b \end{array}\right]$, show that $c_A(x) = (x-a-b)(x+a-b)$ and find an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Exercise 8.2.9 Consider $A = \left[\begin{array}{ccc} b & 0 & a \\ 0 & b & 0 \\ a & 0 & b \end{array}\right]$. Show that $c_A(x) = (x-b)(x-b-a)(x-b+a)$ and find an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Exercise 8.2.10 In each case find new variables y_1 and y_2 that diagonalize the quadratic form q.

a. $q = x_1^2 + 6x_1x_2 + x_2^2$

b. $q = x_1^2 + 4x_1x_2 - 2x_2^2$

Exercise 8.2.11 Show that the following are equivalent for a symmetric matrix A.

a. A is orthogonal.

b. $A^2 = I$.

c. All eigenvalues of A are ± 1.

[Hint: For (b) if and only if (c), use Theorem 8.2.2.]

Exercise 8.2.12 We call matrices A and B **orthogonally similar** (and write $A \overset{\circ}{\sim} B$) if $B = P^TAP$ for an orthogonal matrix P.

a. Show that $A \overset{\circ}{\sim} A$ for all A; $A \overset{\circ}{\sim} B \Rightarrow B \overset{\circ}{\sim} A$; and $A \overset{\circ}{\sim} B$ and $B \overset{\circ}{\sim} C \Rightarrow A \overset{\circ}{\sim} C$.

b. Show that the following are equivalent for two symmetric matrices A and B.

   i. A and B are similar.

   ii. A and B are orthogonally similar.

   iii. A and B have the same eigenvalues.

Exercise 8.2.13 Assume that A and B are orthogonally similar (Exercise 8.2.12).

a. If A and B are invertible, show that A^{-1} and B^{-1} are orthogonally similar.

b. Show that A^2 and B^2 are orthogonally similar.

c. Show that, if A is symmetric, so is B.

Exercise 8.2.14 If A is symmetric, show that every eigenvalue of A is nonnegative if and only if $A = B^2$ for some symmetric matrix B.

Exercise 8.2.15 Prove the converse of Theorem 8.2.3:

If $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$ for all n-columns $\mathbf{x}$ and $\mathbf{y}$, then A is symmetric.

Exercise 8.2.16 Show that every eigenvalue of A is zero if and only if A is nilpotent ($A^k = 0$ for some $k \geq 1$).

Exercise 8.2.17 If A has real eigenvalues, show that $A = B + C$ where B is symmetric and C is nilpotent. [*Hint*: Theorem 8.2.5.]

Exercise 8.2.18 Let P be an orthogonal matrix.

a. Show that $\det P = 1$ or $\det P = -1$.

b. Give 2×2 examples of P such that $\det P = 1$ and $\det P = -1$.

c. If $\det P = -1$, show that $I + P$ has no inverse. [*Hint*: $P^T(I+P) = (I+P)^T$.]

d. If P is $n \times n$ and $\det P \neq (-1)^n$, show that $I - P$ has no inverse.

[*Hint*: $P^T(I-P) = -(I-P)^T$.]

Exercise 8.2.19 We call a square matrix E a **projection matrix** if $E^2 = E = E^T$. (See Exercise 8.1.17.)

a. If E is a projection matrix, show that $P = I - 2E$ is orthogonal and symmetric.

b. If P is orthogonal and symmetric, show that $E = \frac{1}{2}(I - P)$ is a projection matrix.

c. If U is $m \times n$ and $U^TU = I$ (for example, a unit column in $\mathbb{R}^n$), show that $E = UU^T$ is a projection matrix.

Exercise 8.2.20 A matrix that we obtain from the identity matrix by writing its rows in a different order is called a **permutation matrix**. Show that every permutation matrix is orthogonal.

Exercise 8.2.21 If the rows $\mathbf{r}_1, \dots, \mathbf{r}_n$ of the $n \times n$ matrix $A = [a_{ij}]$ are orthogonal, show that the (i, j)-entry of A^{-1} is $\frac{a_{ji}}{\|\mathbf{r}_j\|^2}$.

Exercise 8.2.22

a. Let A be an $m \times n$ matrix. Show that the following are equivalent.

   i. A has orthogonal rows.

   ii. A can be factored as $A = DP$, where D is invertible and diagonal and P has orthonormal rows.

   iii. AA^T is an invertible, diagonal matrix.

b. Show that an $n \times n$ matrix A has orthogonal rows if and only if A can be factored as $A = DP$, where P is orthogonal and D is diagonal and invertible.

Exercise 8.2.23 Let A be a skew-symmetric matrix; that is, $A^T = -A$. Assume that A is an $n \times n$ matrix.

a. Show that $I + A$ is invertible. [*Hint*: By Theorem 2.4.5, it suffices to show that $(I + A)\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, implies $\mathbf{x} = \mathbf{0}$. Compute $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T\mathbf{x}$, and use the fact that $A\mathbf{x} = -\mathbf{x}$ and $A^2\mathbf{x} = \mathbf{x}$.]

b. Show that $P = (I - A)(I + A)^{-1}$ is orthogonal.

c. Show that every orthogonal matrix P such that $I + P$ is invertible arises as in part (b) from some skew-symmetric matrix A.
[*Hint*: Solve $P = (I - A)(I + A)^{-1}$ for A.]

Exercise 8.2.24 Show that the following are equivalent for an $n \times n$ matrix P.

a. P is orthogonal.

b. $\|P\mathbf{x}\| = \|\mathbf{x}\|$ for all columns $\mathbf{x}$ in $\mathbb{R}^n$.

c. $\|P\mathbf{x} - P\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\|$ for all columns $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

d. $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all columns $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

[*Hints*: For (c) $\Rightarrow$ (d), see Exercise 5.3.14(a). For (d) $\Rightarrow$ (a), show that column i of P equals $P\mathbf{e}_i$, where $\mathbf{e}_i$ is column i of the identity matrix.]

Exercise 8.2.25 Show that every 2×2 orthogonal matrix has the form $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ or $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ for some angle θ.

[*Hint*: If $a^2 + b^2 = 1$, then $a = \cos\theta$ and $b = \sin\theta$ for some angle θ.]

Exercise 8.2.26 Use Theorem 8.2.5 to show that every symmetric matrix is orthogonally diagonalizable.

8.3 Positive Definite Matrices

All the eigenvalues of any symmetric matrix are real; this section is about the case in which the eigenvalues are positive. These matrices, which arise whenever optimization (maximum and minimum) problems are encountered, have countless applications throughout science and engineering. They also arise in statistics (for example, in factor analysis used in the social sciences) and in geometry (see Section 8.9). We will encounter them again in Chapter 10 when describing all inner products in $\mathbb{R}^n$.

Definition 8.5 Positive Definite Matrices

A square matrix is called ***positive definite*** *if it is symmetric and all its eigenvalues λ are positive, that is $\lambda > 0$.*

Because these matrices are symmetric, the principal axes theorem plays a central role in the theory.

Theorem 8.3.1

If A is positive definite, then it is invertible and $\det A > 0$.

Proof. If A is $n \times n$ and the eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_n$, then $\det A = \lambda_1 \lambda_2 \cdots \lambda_n > 0$ by the principal axes theorem (or the corollary to Theorem 8.2.5). □

If $\mathbf{x}$ is a column in $\mathbb{R}^n$ and A is any real $n \times n$ matrix, we view the 1×1 matrix $\mathbf{x}^T A\mathbf{x}$ as a real number. With this convention, we have the following characterization of positive definite matrices.

Theorem 8.3.2

A symmetric matrix A is positive definite if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for every column $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$.

Proof. A is symmetric so, by the principal axes theorem, let $P^T AP = D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ where $P^{-1} = P^T$ and the λ_i are the eigenvalues of A. Given a column $\mathbf{x}$ in $\mathbb{R}^n$, write $\mathbf{y} = P^T\mathbf{x} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$. Then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T (PDP^T)\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{8.3}$$

If A is positive definite and $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{x}^T A\mathbf{x} > 0$ by (8.3) because some $y_j \neq 0$ and every $\lambda_i > 0$. Conversely, if $\mathbf{x}^T A\mathbf{x} > 0$ whenever $\mathbf{x} \neq \mathbf{0}$, let $\mathbf{x} = P\mathbf{e}_j \neq \mathbf{0}$ where $\mathbf{e}_j$ is column j of I_n. Then $\mathbf{y} = \mathbf{e}_j$, so (8.3) reads $\lambda_j = \mathbf{x}^T A\mathbf{x} > 0$. □