lyryx with Open Texts

LINEAR ALGEBRA with Applications

Open Edition

Adapted for

Emory University

Math 221

Linear Algebra

Sections 1 & 2

Lectured and adapted by

Le Chen

April 15, 2021

le.chen@emory.edu

Course page

http://math.emory.edu/~lchen41/teaching/2021_Spring_Math221

ADAPTABLE | ACCESSIBLE | AFFORDABLE

by W. Keith Nicholson

Contents

8. Orthogonality

Contents

In Section 5.3 we introduced the dot product in $\mathbb{R}^n$ and extended the basic geometric notions of length and distance. A set $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ of nonzero vectors in $\mathbb{R}^n$ was called an **orthogonal set** if $\mathbf{f}_i \cdot \mathbf{f}_j = 0$ for all $i \neq j$, and it was proved that every orthogonal set is independent. In particular, it was observed that the expansion of a vector as a linear combination of orthogonal basis vectors is easy to obtain because formulas exist for the coefficients. Hence the orthogonal bases are the "nice" bases, and much of this chapter is devoted to extending results about bases to orthogonal bases. This leads to some very powerful methods and theorems. Our first task is to show that every subspace of $\mathbb{R}^n$ *has* an orthogonal basis.

8.1 Orthogonal Complements and Projections

If $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is linearly independent in a general vector space, and if $\mathbf{v}_{m+1}$ is not in $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{v}_{m+1}\}$ is independent (Lemma 6.4.1). Here is the analog for *orthogonal* sets in $\mathbb{R}^n$.

Lemma 8.1.1: Orthogonal Lemma

Let $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ be an orthogonal set in $\mathbb{R}^n$. Given $\mathbf{x}$ in $\mathbb{R}^n$, write

$$\mathbf{f}_{m+1} = \mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

Then:

1. $\mathbf{f}_{m+1}\cdot\mathbf{f}_k = 0$ *for* $k = 1, 2, \dots, m$.
2. *If $\mathbf{x}$ is not in $\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, then $\mathbf{f}_{m+1} \neq \mathbf{0}$ and $\{\mathbf{f}_1, \dots, \mathbf{f}_m, \mathbf{f}_{m+1}\}$ is an orthogonal set.*

Proof. For convenience, write $t_i = (\mathbf{x}\cdot\mathbf{f}_i)/\|\mathbf{f}_i\|^2$ for each i. Given $1 \leq k \leq m$:

$$\begin{aligned}
\mathbf{f}_{m+1}\cdot\mathbf{f}_k &= (\mathbf{x} - t_1\mathbf{f}_1 - \cdots - t_k\mathbf{f}_k - \cdots - t_m\mathbf{f}_m)\cdot\mathbf{f}_k \\
&= \mathbf{x}\cdot\mathbf{f}_k - t_1(\mathbf{f}_1\cdot\mathbf{f}_k) - \cdots - t_k(\mathbf{f}_k\cdot\mathbf{f}_k) - \cdots - t_m(\mathbf{f}_m\cdot\mathbf{f}_k) \\
&= \mathbf{x}\cdot\mathbf{f}_k - t_k\|\mathbf{f}_k\|^2 \\
&= 0
\end{aligned}$$

This proves (1), and (2) follows because $\mathbf{f}_{m+1} \neq \mathbf{0}$ if $\mathbf{x}$ is not in $\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$. □

The orthogonal lemma has three important consequences for $\mathbb{R}^n$. The first is an extension for orthogonal sets of the fundamental fact that any independent set is part of a basis (Theorem 6.4.1).

Theorem 8.1.1

Let U be a subspace of $\mathbb{R}^n$.

1. *Every orthogonal subset $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ in U is a subset of an orthogonal basis of U.*
2. *U has an orthogonal basis.*

Proof.

1. If $\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m\} = U$, it is *already* a basis. Otherwise, there exists $\mathbf{x}$ in U outside $\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$. If $\mathbf{f}_{m+1}$ is as given in the orthogonal lemma, then $\mathbf{f}_{m+1}$ is in U and $\{\mathbf{f}_1, \dots, \mathbf{f}_m, \mathbf{f}_{m+1}\}$ is orthogonal. If $\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m, \mathbf{f}_{m+1}\} = U$, we are done. Otherwise,

the process continues to create larger and larger orthogonal subsets of U. They are all independent by Theorem 5.3.5, so we have a basis when we reach a subset containing dim U vectors.

2. If $U = \{\mathbf{0}\}$, the empty basis is orthogonal. Otherwise, if $\mathbf{f} \neq \mathbf{0}$ is in U, then $\{\mathbf{f}\}$ is orthogonal, so (2) follows from (1). □

We can improve upon (2) of Theorem 8.1.1. In fact, the second consequence of the orthogonal lemma is a procedure by which *any* basis $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ of a subspace U of $\mathbb{R}^n$ can be systematically modified to yield an orthogonal basis $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ of U. The $\mathbf{f}_i$ are constructed one at a time from the $\mathbf{x}_i$.

To start the process, take $\mathbf{f}_1 = \mathbf{x}_1$. Then $\mathbf{x}_2$ is not in $\text{span}\{\mathbf{f}_1\}$ because $\{\mathbf{x}_1, \mathbf{x}_2\}$ is independent, so take

$$\mathbf{f}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1$$

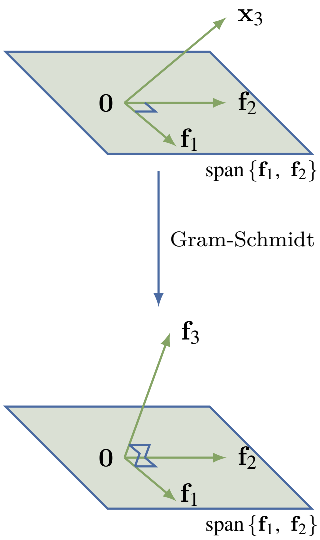
Thus $\{\mathbf{f}_1, \mathbf{f}_2\}$ is orthogonal by Lemma 8.1.1. Moreover, $\text{span}\{\mathbf{f}_1, \mathbf{f}_2\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$ (verify), so $\mathbf{x}_3$ is not in $\text{span}\{\mathbf{f}_1, \mathbf{f}_2\}$. Hence $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ is orthogonal where

$$\mathbf{f}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2$$

Again, $\text{span}\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, so $\mathbf{x}_4$ is not in $\text{span}\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ and the process continues. At the mth iteration we construct an orthogonal set $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ such that

$$\text{span}\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\} = U$$

Hence $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ is the desired orthogonal basis of U. The procedure can be summarized as follows.



Theorem 8.1.2: Gram-Schmidt Orthogonalization Algorithm[1]

If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ is any basis of a subspace U of $\mathbb{R}^n$, construct $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m$ in U successively as follows:

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{x}_1 \\
\mathbf{f}_2 &= \mathbf{x}_2 - \tfrac{\mathbf{x}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 \\
\mathbf{f}_3 &= \mathbf{x}_3 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 \\
&\vdots \\
\mathbf{f}_k &= \mathbf{x}_k - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_{k-1}}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1}
\end{aligned}$$

for each $k = 2, 3, \dots, m$. Then

1. *$\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ is an orthogonal basis of U.*
2. *$\operatorname{span}\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k\} = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ for each $k = 1, 2, \dots, m$.*

The process (for $k = 3$) is depicted in the diagrams. Of course, the algorithm converts any basis of $\mathbb{R}^n$ itself into an orthogonal basis.

Example 8.1.1

Find an orthogonal basis of the row space of $A = \left[\begin{array}{rrrr} 1 & 1 & -1 & -1 \\ 3 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right]$.

Solution. Let $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ denote the rows of A and observe that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is linearly independent. Take $\mathbf{f}_1 = \mathbf{x}_1$. The algorithm gives

$$\begin{aligned}
\mathbf{f}_2 &= \mathbf{x}_2 - \tfrac{\mathbf{x}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 = (3, 2, 0, 1) - \tfrac{4}{4}(1, 1, -1, -1) = (2, 1, 1, 2) \\
\mathbf{f}_3 &= \mathbf{x}_3 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 = \mathbf{x}_3 - \tfrac{0}{4}\mathbf{f}_1 - \tfrac{3}{10}\mathbf{f}_2 = \tfrac{1}{10}(4, -3, 7, -6)
\end{aligned}$$

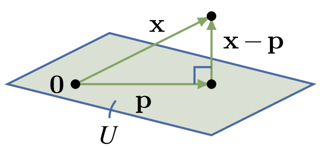
Hence $\{(1, 1, -1, -1), (2, 1, 1, 2), \frac{1}{10}(4, -3, 7, -6)\}$ is the orthogonal basis provided by the algorithm. In hand calculations it may be convenient to eliminate fractions (see the Remark below), so $\{(1, 1, -1, -1), (2, 1, 1, 2), (4, -3, 7, -6)\}$ is also an orthogonal basis for row A.

[1] Erhardt Schmidt (1876–1959) was a German mathematician who studied under the great David Hilbert and later developed the theory of Hilbert spaces. He first described the present algorithm in 1907. Jörgen Pederson Gram (1850–1916) was a Danish actuary.

Remark

Observe that the vector $\frac{\mathbf{x} \cdot \mathbf{f}_i}{\|\mathbf{f}_i\|^2}\mathbf{f}_i$ is unchanged if a nonzero scalar multiple of $\mathbf{f}_i$ is used in place of $\mathbf{f}_i$. Hence, if a newly constructed $\mathbf{f}_i$ is multiplied by a nonzero scalar at some stage of the Gram-Schmidt algorithm, the subsequent $\mathbf{f}$s will be unchanged. This is useful in actual calculations.

Projections



Suppose a point $\mathbf{x}$ and a plane U through the origin in $\mathbb{R}^3$ are given, and we want to find the point $\mathbf{p}$ in the plane that is closest to $\mathbf{x}$. Our geometric intuition assures us that such a point $\mathbf{p}$ exists. In fact (see the diagram), $\mathbf{p}$ must be chosen in such a way that $\mathbf{x} - \mathbf{p}$ is *perpendicular* to the plane.

Now we make two observations: first, the plane U is a *subspace of* $\mathbb{R}^3$ (because U contains the origin); and second, that the condition that $\mathbf{x} - \mathbf{p}$ is perpendicular to the plane U means that $\mathbf{x} - \mathbf{p}$ is *orthogonal* to every vector in U. In these terms the whole discussion makes sense in $\mathbb{R}^n$. Furthermore, the orthogonal lemma provides exactly what is needed to find $\mathbf{p}$ in this more general setting.

Definition 8.1 Orthogonal Complement of a Subspace of $\mathbb{R}^n$

*If U is a subspace of $\mathbb{R}^n$, define the **orthogonal complement** $U^\perp$ of U (pronounced "U-perp") by*

$$U^\perp = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \text{ in } U\}$$

The following lemma collects some useful properties of the orthogonal complement; the proof of (1) and (2) is left as Exercise 8.1.6.

Lemma 8.1.2

Let U be a subspace of $\mathbb{R}^n$.

1. *$U^\perp$ is a subspace of $\mathbb{R}^n$.*
2. *$\{\mathbf{0}\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$.*
3. *If $U = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, then $U^\perp = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{x}_i = 0 \text{ for } i = 1, 2, \dots, k\}$.*

Proof.

3. Let $U = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$; we must show that $U^\perp = \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{x}_i = 0 \text{ for each } i\}$. If $\mathbf{x}$ is in $U^\perp$ then $\mathbf{x} \cdot \mathbf{x}_i = 0$ for all i because each $\mathbf{x}_i$ is in U. Conversely, suppose that $\mathbf{x} \cdot \mathbf{x}_i = 0$ for all i; we must show that $\mathbf{x}$ is in $U^\perp$, that is, $\mathbf{x} \cdot \mathbf{y} = 0$ for each $\mathbf{y}$ in U. Write $\mathbf{y} = r_1\mathbf{x}_1 + r_2\mathbf{x}_2 + \cdots + r_k\mathbf{x}_k$, where each r_i is in $\mathbb{R}$. Then, using Theorem 5.3.1,

$$\mathbf{x} \cdot \mathbf{y} = r_1(\mathbf{x} \cdot \mathbf{x}_1) + r_2(\mathbf{x} \cdot \mathbf{x}_2) + \cdots + r_k(\mathbf{x} \cdot \mathbf{x}_k) = r_1 0 + r_2 0 + \cdots + r_k 0 = 0$$

as required. □

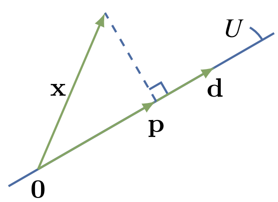
Example 8.1.2

Find $U^{\perp}$ if $U = \text{span}\{(1,\ -1,\ 2,\ 0),\ (1,\ 0,\ -2,\ 3)\}$ in $\mathbb{R}^4$.

Solution. By Lemma 8.1.2, $\mathbf{x} = (x,\ y,\ z,\ w)$ is in $U^{\perp}$ if and only if it is orthogonal to both $(1,\ -1,\ 2,\ 0)$ and $(1,\ 0,\ -2,\ 3)$; that is,

$$\begin{aligned} x - y + 2z \phantom{{}+3w} &= 0 \\ x \phantom{{}-y} - 2z + 3w &= 0 \end{aligned}$$

Gaussian elimination gives $U^{\perp} = \text{span}\{(2,\ 4,\ 1,\ 0),\ (3,\ 3,\ 0,\ -1)\}$.



Now consider vectors $\mathbf{x}$ and $\mathbf{d} \neq \mathbf{0}$ in $\mathbb{R}^3$. The projection $\mathbf{p} = \text{proj}_{\mathbf{d}}\, \mathbf{x}$ of $\mathbf{x}$ on $\mathbf{d}$ was defined in Section 4.2 as in the diagram.

The following formula for $\mathbf{p}$ was derived in Theorem 4.2.4

$$\mathbf{p} = \text{proj}_{\mathbf{d}}\, \mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{d}}{\|\mathbf{d}\|^2}\right)\mathbf{d}$$

where it is shown that $\mathbf{x} - \mathbf{p}$ is orthogonal to $\mathbf{d}$. Now observe that the line $U = \mathbb{R}\mathbf{d} = \{t\mathbf{d} \mid t \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$, that $\{\mathbf{d}\}$ is an orthogonal basis of U, and that $\mathbf{p} \in U$ and $\mathbf{x} - \mathbf{p} \in U^{\perp}$ (by Theorem 4.2.4).

In this form, this makes sense for any vector $\mathbf{x}$ in $\mathbb{R}^n$ and any subspace U of $\mathbb{R}^n$, so we generalize it as follows. If $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \dots,\ \mathbf{f}_m\}$ is an orthogonal basis of U, we define the projection $\mathbf{p}$ of $\mathbf{x}$ on U by the formula

$$\mathbf{p} = \left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right)\mathbf{f}_1 + \left(\frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right)\mathbf{f}_2 + \cdots + \left(\frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\right)\mathbf{f}_m \tag{8.1}$$

Then $\mathbf{p} \in U$ and (by the orthogonal lemma) $\mathbf{x} - \mathbf{p} \in U^{\perp}$, so it looks like we have a generalization of Theorem 4.2.4.

However there is a potential problem: the formula (8.1) for $\mathbf{p}$ must be shown to be independent of the choice of the orthogonal basis $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \dots,\ \mathbf{f}_m\}$. To verify this, suppose that $\{\mathbf{f}'_1,\ \mathbf{f}'_2,\ \dots,\ \mathbf{f}'_m\}$ is another orthogonal basis of U, and write

$$\mathbf{p}' = \left(\frac{\mathbf{x}\cdot\mathbf{f}'_1}{\|\mathbf{f}'_1\|^2}\right)\mathbf{f}'_1 + \left(\frac{\mathbf{x}\cdot\mathbf{f}'_2}{\|\mathbf{f}'_2\|^2}\right)\mathbf{f}'_2 + \cdots + \left(\frac{\mathbf{x}\cdot\mathbf{f}'_m}{\|\mathbf{f}'_m\|^2}\right)\mathbf{f}'_m$$

As before, $\mathbf{p}' \in U$ and $\mathbf{x} - \mathbf{p}' \in U^{\perp}$, and we must show that $\mathbf{p}' = \mathbf{p}$. To see this, write the vector $\mathbf{p} - \mathbf{p}'$ as follows:

$$\mathbf{p} - \mathbf{p}' = (\mathbf{x} - \mathbf{p}') - (\mathbf{x} - \mathbf{p})$$

This vector is in U (because $\mathbf{p}$ and $\mathbf{p}'$ are in U) and it is in $U^{\perp}$ (because $\mathbf{x} - \mathbf{p}'$ and $\mathbf{x} - \mathbf{p}$ are in $U^{\perp}$), and so it must be zero (it is orthogonal to itself!). This means $\mathbf{p}' = \mathbf{p}$ as desired.

Hence, the vector $\mathbf{p}$ in equation (8.1) depends only on $\mathbf{x}$ and the subspace U, and *not* on the choice of orthogonal basis $\{\mathbf{f}_1,\ \dots,\ \mathbf{f}_m\}$ of U used to compute it. Thus, we are entitled to make the following definition:

Definition 8.2 Projection onto a Subspace of $\mathbb{R}^n$

Let U be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$. If $\mathbf{x}$ is in $\mathbb{R}^n$, the vector

$$\text{proj}_U \mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \dots + \frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

*is called the **orthogonal projection** of $\mathbf{x}$ on U. For the zero subspace $U = \{\mathbf{0}\}$, we define*

$$\text{proj}_{\{\mathbf{0}\}} \mathbf{x} = \mathbf{0}$$

The preceding discussion proves (1) of the following theorem.

Theorem 8.1.3: Projection Theorem

If U is a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ is in $\mathbb{R}^n$, write $\mathbf{p} = \text{proj}_U \mathbf{x}$. Then:

1. *$\mathbf{p}$ is in U and $\mathbf{x} - \mathbf{p}$ is in $U^\perp$.*
2. *$\mathbf{p}$ is the vector in U closest to $\mathbf{x}$ in the sense that*

$$\|\mathbf{x} - \mathbf{p}\| < \|\mathbf{x} - \mathbf{y}\| \quad \text{for all } \mathbf{y} \in U,\ \mathbf{y} \neq \mathbf{p}$$

Proof.

1. This is proved in the preceding discussion (it is clear if $U = \{\mathbf{0}\}$).
2. Write $\mathbf{x} - \mathbf{y} = (\mathbf{x} - \mathbf{p}) + (\mathbf{p} - \mathbf{y})$. Then $\mathbf{p} - \mathbf{y}$ is in U and so is orthogonal to $\mathbf{x} - \mathbf{p}$ by (1). Hence, the Pythagorean theorem gives

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{p} - \mathbf{y}\|^2 > \|\mathbf{x} - \mathbf{p}\|^2$$

because $\mathbf{p} - \mathbf{y} \neq \mathbf{0}$. This gives (2). □

Example 8.1.3

Let $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$ in $\mathbb{R}^4$ where $\mathbf{x}_1 = (1, 1, 0, 1)$ and $\mathbf{x}_2 = (0, 1, 1, 2)$. If $\mathbf{x} = (3, -1, 0, 2)$, find the vector in U closest to $\mathbf{x}$ and express $\mathbf{x}$ as the sum of a vector in U and a vector orthogonal to U.

Solution. $\{\mathbf{x}_1, \mathbf{x}_2\}$ is independent but not orthogonal. The Gram-Schmidt process gives an orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ of U where $\mathbf{f}_1 = \mathbf{x}_1 = (1, 1, 0, 1)$ and

$$\mathbf{f}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 = \mathbf{x}_2 - \frac{3}{3}\mathbf{f}_1 = (-1, 0, 1, 1)$$

Hence, we can compute the projection using $\{\mathbf{f}_1, \mathbf{f}_2\}$:

$$\mathbf{p} = \text{proj}_U \mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 = \frac{4}{3}\mathbf{f}_1 + \frac{-1}{3}\mathbf{f}_2 = \frac{1}{3}\begin{bmatrix} 5 & 4 & -1 & 3 \end{bmatrix}$$

Thus, $\mathbf{p}$ is the vector in U closest to $\mathbf{x}$, and $\mathbf{x}-\mathbf{p} = \frac{1}{3}(4,\ -7,\ 1,\ 3)$ is orthogonal to every vector in U. (This can be verified by checking that it is orthogonal to the generators $\mathbf{x}_1$ and $\mathbf{x}_2$ of U.) The required decomposition of $\mathbf{x}$ is thus

$$\mathbf{x} = \mathbf{p} + (\mathbf{x}-\mathbf{p}) = \tfrac{1}{3}(5,\ 4,\ -1,\ 3) + \tfrac{1}{3}(4,\ -7,\ 1,\ 3)$$

Example 8.1.4

Find the point in the plane with equation $2x+y-z=0$ that is closest to the point $(2,\ -1,\ -3)$.

Solution. We write $\mathbb{R}^3$ as rows. The plane is the subspace U whose points $(x,\ y,\ z)$ satisfy $z = 2x+y$. Hence

$$U = \{(s,\ t,\ 2s+t) \mid s,\ t \text{ in } \mathbb{R}\} = \operatorname{span}\{(0,\ 1,\ 1),\ (1,\ 0,\ 2)\}$$

The Gram-Schmidt process produces an orthogonal basis $\{\mathbf{f}_1,\ \mathbf{f}_2\}$ of U where $\mathbf{f}_1 = (0,\ 1,\ 1)$ and $\mathbf{f}_2 = (1,\ -1,\ 1)$. Hence, the vector in U closest to $\mathbf{x} = (2,\ -1,\ -3)$ is

$$\operatorname{proj}_U \mathbf{x} = \tfrac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \tfrac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 = -2\mathbf{f}_1 + 0\mathbf{f}_2 = (0,\ -2,\ -2)$$

Thus, the point in U closest to $(2,\ -1,\ -3)$ is $(0,\ -2,\ -2)$.

The next theorem shows that projection on a subspace of $\mathbb{R}^n$ is actually a linear operator $\mathbb{R}^n \to \mathbb{R}^n$.

Theorem 8.1.4

Let U be a fixed subspace of $\mathbb{R}^n$. If we define $T : \mathbb{R}^n \to \mathbb{R}^n$ by

$$T(\mathbf{x}) = \operatorname{proj}_U \mathbf{x} \quad \textit{for all } \mathbf{x} \textit{ in } \mathbb{R}^n$$

1. *T is a linear operator.*
2. *$\operatorname{im} T = U$ and $\ker T = U^\perp$.*
3. *$\dim U + \dim U^\perp = n$.*

Proof. If $U = \{\mathbf{0}\}$, then $U^\perp = \mathbb{R}^n$, and so $T(\mathbf{x}) = \operatorname{proj}_{\{\mathbf{0}\}} \mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$. Thus $T = 0$ is the zero (linear) operator, so (1), (2), and (3) hold. Hence assume that $U \neq \{\mathbf{0}\}$.

1. If $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \dots,\ \mathbf{f}_m\}$ is an orthonormal basis of U, then

$$T(\mathbf{x}) = (\mathbf{x}\cdot\mathbf{f}_1)\mathbf{f}_1 + (\mathbf{x}\cdot\mathbf{f}_2)\mathbf{f}_2 + \cdots + (\mathbf{x}\cdot\mathbf{f}_m)\mathbf{f}_m \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{8.2}$$

   by the definition of the projection. Thus T is linear because

$$(\mathbf{x}+\mathbf{y})\cdot\mathbf{f}_i = \mathbf{x}\cdot\mathbf{f}_i + \mathbf{y}\cdot\mathbf{f}_i \quad \text{and} \quad (r\mathbf{x})\cdot\mathbf{f}_i = r(\mathbf{x}\cdot\mathbf{f}_i) \quad \text{for each } i$$

2. We have $\operatorname{im} T \subseteq U$ by (8.2) because each $\mathbf{f}_i$ is in U. But if $\mathbf{x}$ is in U, then $\mathbf{x} = T(\mathbf{x})$ by (8.2) and the expansion theorem applied to the space U. This shows that $U \subseteq \operatorname{im} T$, so $\operatorname{im} T = U$. Now suppose that $\mathbf{x}$ is in $U^\perp$. Then $\mathbf{x} \cdot \mathbf{f}_i = 0$ for each i (again because each $\mathbf{f}_i$ is in U) so $\mathbf{x}$ is in $\ker T$ by (8.2). Hence $U^\perp \subseteq \ker T$. On the other hand, Theorem 8.1.3 shows that $\mathbf{x} - T(\mathbf{x})$ is in $U^\perp$ for all $\mathbf{x}$ in $\mathbb{R}^n$, and it follows that $\ker T \subseteq U^\perp$. Hence $\ker T = U^\perp$, proving (2).

3. This follows from (1), (2), and the dimension theorem (Theorem 7.2.4).

□

Exercises for 8.1

Exercise 8.1.1 In each case, use the Gram-Schmidt algorithm to convert the given basis B of V into an orthogonal basis.

a. $V = \mathbb{R}^2$, $B = \{(1, -1), (2, 1)\}$

b. $V = \mathbb{R}^2$, $B = \{(2, 1), (1, 2)\}$

c. $V = \mathbb{R}^3$, $B = \{(1, -1, 1), (1, 0, 1), (1, 1, 2)\}$

d. $V = \mathbb{R}^3$, $B = \{(0, 1, 1), (1, 1, 1), (1, -2, 2)\}$

---

b. $\{(2, 1), \frac{3}{5}(-1, 2)\}$

d. $\{(0, 1, 1), (1, 0, 0), (0, -2, 2)\}$

Exercise 8.1.2 In each case, write $\mathbf{x}$ as the sum of a vector in U and a vector in $U^\perp$.

a. $\mathbf{x} = (1, 5, 7)$, $U = \text{span}\{(1, -2, 3), (-1, 1, 1)\}$

b. $\mathbf{x} = (2, 1, 6)$, $U = \text{span}\{(3, -1, 2), (2, 0, -3)\}$

c. $\mathbf{x} = (3, 1, 5, 9)$,
$U = \text{span}\{(1, 0, 1, 1), (0, 1, -1, 1), (-2, 0, 1, 1)\}$

d. $\mathbf{x} = (2, 0, 1, 6)$,
$U = \text{span}\{(1, 1, 1, 1), (1, 1, -1, -1), (1, -1, 1, -1)\}$

e. $\mathbf{x} = (a, b, c, d)$,
$U = \text{span}\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)\}$

f. $\mathbf{x} = (a, b, c, d)$,
$U = \text{span}\{(1, -1, 2, 0), (-1, 1, 1, 1)\}$

---

b. $\mathbf{x} = \frac{1}{182}(271, -221, 1030) + \frac{1}{182}(93, 403, 62)$

d. $\mathbf{x} = \frac{1}{4}(1, 7, 11, 17) + \frac{1}{4}(7, -7, -7, 7)$

f. $\mathbf{x} = \frac{1}{12}(5a - 5b + c - 3d, -5a + 5b - c + 3d, a - b + 11c + 3d, -3a + 3b + 3c + 3d) + \frac{1}{12}(7a + 5b - c + 3d, 5a + 7b + c - 3d, -a + b + c - 3d, 3a - 3b - 3c + 9d)$

Exercise 8.1.3 Let $\mathbf{x} = (1, -2, 1, 6)$ in $\mathbb{R}^4$, and let $U = \text{span}\{(2, 1, 3, -4), (1, 2, 0, 1)\}$.

a. Compute $\text{proj}_U \mathbf{x}$.

b. Show that $\{(1, 0, 2, -3), (4, 7, 1, 2)\}$ is another orthogonal basis of U.

c. Use the basis in part (b) to compute $\text{proj}_U \mathbf{x}$.

---

a. $\frac{1}{10}(-9, 3, -21, 33) = \frac{3}{10}(-3, 1, -7, 11)$

c. $\frac{1}{70}(-63, 21, -147, 231) = \frac{3}{10}(-3, 1, -7, 11)$

Exercise 8.1.4 In each case, use the Gram-Schmidt algorithm to find an orthogonal basis of the subspace U, and find the vector in U closest to $\mathbf{x}$.

a. $U = \text{span}\{(1, 1, 1), (0, 1, 1)\}$, $\mathbf{x} = (-1, 2, 1)$

b. $U = \text{span}\{(1, -1, 0), (-1, 0, 1)\}$, $\mathbf{x} = (2, 1, 0)$

c. $U = \text{span}\{(1, 0, 1, 0), (1, 1, 1, 0), (1, 1, 0, 0)\}$,
$\mathbf{x} = (2, 0, -1, 3)$

d. $U = \text{span}\{(1, -1, 0, 1), (1, 1, 0, 0), (1, 1, 0, 1)\}$, $\mathbf{x} = (2, 0, 3, 1)$

---

b. $\{(1, -1, 0), \frac{1}{2}(-1, -1, 2)\}$; $\text{proj}_U \mathbf{x} = (1, 0, -1)$

d. $\{(1, -1, 0, 1), (1, 1, 0, 0), \frac{1}{3}(-1, 1, 0, 2)\}$; $\text{proj}_U \mathbf{x} = (2, 0, 0, 1)$

Exercise 8.1.5 Let $U = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, $\mathbf{v}_i$ in $\mathbb{R}^n$, and let A be the $k \times n$ matrix with the $\mathbf{v}_i$ as rows.

a. Show that $U^\perp = \{\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n, A\mathbf{x}^T = \mathbf{0}\}$.

b. Use part (a) to find $U^\perp$ if $U = \text{span}\{(1, -1, 2, 1), (1, 0, -1, 1)\}$.

---

b. $U^\perp = \text{span}\{(1, 3, 1, 0), (-1, 0, 0, 1)\}$

Exercise 8.1.6

a. Prove part 1 of Lemma 8.1.2.

b. Prove part 2 of Lemma 8.1.2.

Exercise 8.1.7 Let U be a subspace of $\mathbb{R}^n$. If $\mathbf{x}$ in $\mathbb{R}^n$ can be written in any way at all as $\mathbf{x} = \mathbf{p} + \mathbf{q}$ with $\mathbf{p}$ in U and $\mathbf{q}$ in $U^\perp$, show that necessarily $\mathbf{p} = \text{proj}_U \mathbf{x}$.

Exercise 8.1.8 Let U be a subspace of $\mathbb{R}^n$ and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$. Using Exercise 8.1.7, or otherwise, show that $\mathbf{x}$ is in U if and only if $\mathbf{x} = \text{proj}_U \mathbf{x}$.

Write $\mathbf{p} = \text{proj}_U \mathbf{x}$. Then $\mathbf{p}$ is in U by definition. If $\mathbf{x}$ is U, then $\mathbf{x} - \mathbf{p}$ is in U. But $\mathbf{x} - \mathbf{p}$ is also in $U^\perp$ by Theorem 8.1.3, so $\mathbf{x} - \mathbf{p}$ is in $U \cap U^\perp = \{\mathbf{0}\}$. Thus $\mathbf{x} = \mathbf{p}$.

Exercise 8.1.9 Let U be a subspace of $\mathbb{R}^n$.

a. Show that $U^\perp = \mathbb{R}^n$ if and only if $U = \{\mathbf{0}\}$.

b. Show that $U^\perp = \{\mathbf{0}\}$ if and only if $U = \mathbb{R}^n$.

Exercise 8.1.10 If U is a subspace of $\mathbb{R}^n$, show that $\text{proj}_U \mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in U.

Let $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ be an orthonormal basis of U. If $\mathbf{x}$ is in U the expansion theorem gives $\mathbf{x} = (\mathbf{x} \cdot \mathbf{f}_1)\mathbf{f}_1 + (\mathbf{x} \cdot \mathbf{f}_2)\mathbf{f}_2 + \cdots + (\mathbf{x} \cdot \mathbf{f}_m)\mathbf{f}_m = \text{proj}_U \mathbf{x}$.

Exercise 8.1.11 If U is a subspace of $\mathbb{R}^n$, show that $\mathbf{x} = \text{proj}_U \mathbf{x} + \text{proj}_{U^\perp} \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

Exercise 8.1.12 If $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ is an orthogonal basis of $\mathbb{R}^n$ and $U = \text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, show that $U^\perp = \text{span}\{\mathbf{f}_{m+1}, \dots, \mathbf{f}_n\}$.

Exercise 8.1.13 If U is a subspace of $\mathbb{R}^n$, show that $U^{\perp\perp} = U$. [*Hint*: Show that $U \subseteq U^{\perp\perp}$, then use Theorem 8.1.4 (3) twice.]

Exercise 8.1.14 If U is a subspace of $\mathbb{R}^n$, show how to find an $n \times n$ matrix A such that $U = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$. [*Hint*: Exercise 8.1.13.]

Let $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m\}$ be a basis of $U^\perp$, and let A be the $n \times n$ matrix with rows $\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_m^T, 0, \dots, 0$. Then $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{y}_i \cdot \mathbf{x} = 0$ for each $i = 1, 2, \dots, m$; if and only if $\mathbf{x}$ is in $U^{\perp\perp} = U$.

Exercise 8.1.15 Write $\mathbb{R}^n$ as rows. If A is an $n \times n$ matrix, write its null space as $\text{null}\, A = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x}^T = \mathbf{0}\}$. Show that:

a) $\text{null}\, A = (\text{row}\, A)^\perp$; b) $\text{null}\, A^T = (\text{col}\, A)^\perp$.

Exercise 8.1.16 If U and W are subspaces, show that $(U + W)^\perp = U^\perp \cap W^\perp$. [See Exercise 5.1.22.]

Exercise 8.1.17 Think of $\mathbb{R}^n$ as consisting of rows.

a. Let E be an $n \times n$ matrix, and let $U = \{\mathbf{x}E \mid \mathbf{x} \text{ in } \mathbb{R}^n\}$. Show that the following are equivalent.

   i. $E^2 = E = E^T$ (E is a **projection matrix**).

   ii. $(\mathbf{x} - \mathbf{x}E) \cdot (\mathbf{y}E) = 0$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

   iii. $\text{proj}_U \mathbf{x} = \mathbf{x}E$ for all $\mathbf{x}$ in $\mathbb{R}^n$. [*Hint*: For (ii) implies (iii): Write $\mathbf{x} = \mathbf{x}E + (\mathbf{x} - \mathbf{x}E)$ and use the uniqueness argument preceding the definition of $\text{proj}_U \mathbf{x}$. For (iii) implies (ii): $\mathbf{x} - \mathbf{x}E$ is in $U^\perp$ for all $\mathbf{x}$ in $\mathbb{R}^n$.]

b. If E is a projection matrix, show that $I - E$ is also a projection matrix.

c. If $EF = 0 = FE$ and E and F are projection matrices, show that $E+F$ is also a projection matrix.

d. If A is $m \times n$ and AA^T is invertible, show that $E = A^T(AA^T)^{-1}A$ is a projection matrix.

---

d. $E^T = A^T[(AA^T)^{-1}]^T(A^T)^T = A^T[(AA^T)^T]^{-1}A = A^T[AA^T]^{-1}A = E$ $E^2 = A^T(AA^T)^{-1}AA^T(AA^T)^{-1}A = A^T(AA^T)^{-1}A = E$

Exercise 8.1.18 Let A be an $n \times n$ matrix of rank r. Show that there is an invertible $n \times n$ matrix U such that UA is a row-echelon matrix with the property that the first r rows are orthogonal. [*Hint*: Let R be the row-echelon form of A, and use the Gram-Schmidt process on the nonzero rows of R from the bottom up. Use Lemma 2.4.1.]

Exercise 8.1.19 Let A be an $(n-1) \times n$ matrix with rows $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}$ and let A_i denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting column i. Define the vector $\mathbf{y}$ in $\mathbb{R}^n$ by

$$\mathbf{y} = \left[\begin{array}{ccccc} \det A_1 & -\det A_2 & \det A_3 & \cdots & (-1)^{n+1} \det A_n \end{array} \right]$$

Show that:

a. $\mathbf{x}_i \cdot \mathbf{y} = 0$ for all $i = 1, 2, \dots, n-1$. [*Hint*: Write $B_i = \left[\begin{array}{c} x_i \\ A \end{array} \right]$ and show that $\det B_i = 0$.]

b. $\mathbf{y} \neq \mathbf{0}$ if and only if $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}\}$ is linearly independent. [*Hint*: If some $\det A_i \neq 0$, the rows of A_i are linearly independent. Conversely, if the $\mathbf{x}_i$ are independent, consider $A = UR$ where R is in reduced row-echelon form.]

c. If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}\}$ is linearly independent, use Theorem 8.1.3(3) to show that all solutions to the system of $n-1$ homogeneous equations

$$A\mathbf{x}^T = \mathbf{0}$$

are given by $t\mathbf{y}$, t a parameter.

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 $\left[\begin{array}{rr} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right]$ 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,

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

we will prove (Theorem **??**) 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

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

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 $P^{-1}AP = P^TAP$ 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 **??**. 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 **??**).

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

$$P = \begin{bmatrix} \frac{1}{\sqrt{6}}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{\sqrt{30}}\mathbf{x}_3 \end{bmatrix} = \frac{1}{\sqrt{30}} \begin{bmatrix} \sqrt{5} & 2\sqrt{6} & -1 \\ -2\sqrt{5} & \sqrt{6} & 2 \\ \sqrt{5} & 0 & 5 \end{bmatrix}$$

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

$$P^T A P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

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.

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

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 $\lambda \neq \mu$. □

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 $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

Solution. The characteristic polynomial is

$$c_A(x) = \det \begin{bmatrix} x-8 & 2 & -2 \\ 2 & x-5 & -4 \\ -2 & -4 & x-5 \end{bmatrix} = x(x-9)^2$$

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

$$E_0(A) = \operatorname{span}\{\mathbf{x}_1\}, \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \quad \text{and} \quad E_9(A) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

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

$$\{\mathbf{x}_2,\ \mathbf{x}_3\} \text{ of } E_9(A) \quad \text{where} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{x}_3 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

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

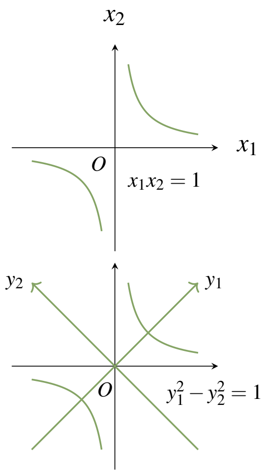
$$P = \begin{bmatrix} \frac{1}{3}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{3\sqrt{5}}\mathbf{x}_3 \end{bmatrix} = \frac{1}{3\sqrt{5}} \begin{bmatrix} \sqrt{5} & -6 & 2 \\ 2\sqrt{5} & 3 & 4 \\ -2\sqrt{5} & 0 & 5 \end{bmatrix}$$

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, $\mathbf{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ and $\mathbf{y}_3 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ 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 $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y}$ 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

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

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

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 $P_1 = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$. Then P_1 is orthogonal and $P_1^T AP_1 = \begin{bmatrix} \lambda_1 & B \\ 0 & A_1 \end{bmatrix}$ 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 $P_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal, so $P = P_1P_2$ is also orthogonal and $P^T AP = \begin{bmatrix} \lambda_1 & BQ \\ 0 & T_1 \end{bmatrix}$ 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 ?? 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 ??).

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:

[4]There is also a lower triangular version.

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* $\operatorname{tr} A = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

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) $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ b) $A = \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$

c) $A = \begin{bmatrix} 1 & 2 \\ -4 & 2 \end{bmatrix}$

d) $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, $(a, b) \neq (0, 0)$

e) $A = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 2 \end{bmatrix}$

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

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

h) $A = \begin{bmatrix} 2 & 6 & -3 \\ 3 & 2 & 6 \\ -6 & 3 & 2 \end{bmatrix}$

---

b. $\frac{1}{5}\begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$

d. $\frac{1}{\sqrt{a^2+b^2}}\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

f. $\begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

h. $\frac{1}{7}\begin{bmatrix} 2 & 6 & -3 \\ 3 & 2 & 6 \\ -6 & 3 & 2 \end{bmatrix}$

Exercise 8.2.2 is diagonal and that all diagonal entries are 1 or -1. ______

We have $P^T = P^{-1}$; this matrix is lower triangular (left side) and also upper triangular (right side–see Lemma 2.7.1), and so is diagonal. But then $P = P^T = P^{-1}$, so $P^2 = I$. This implies that the diagonal entries of P are all ± 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) $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ b) $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

c) $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 5 \end{bmatrix}$ d) $A = \begin{bmatrix} 3 & 0 & 7 \\ 0 & 5 & 0 \\ 7 & 0 & 3 \end{bmatrix}$

e) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ f) $A = \begin{bmatrix} 5 & -2 & -4 \\ -2 & 8 & -2 \\ -4 & -2 & 5 \end{bmatrix}$

g) $A = \begin{bmatrix} 5 & 3 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 1 & 7 \end{bmatrix}$

h) $A = \begin{bmatrix} 3 & 5 & -1 & 1 \\ 5 & 3 & 1 & -1 \\ -1 & 1 & 3 & 5 \\ 1 & -1 & 5 & 3 \end{bmatrix}$

---

b. $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

d. $\frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}$

f. $\frac{1}{3\sqrt{2}} \begin{bmatrix} 2\sqrt{2} & 3 & 1 \\ \sqrt{2} & 0 & -4 \\ 2\sqrt{2} & -3 & 1 \end{bmatrix}$ or $\frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}$

h. $\frac{1}{2} \begin{bmatrix} 1 & -1 & \sqrt{2} & 0 \\ -1 & 1 & \sqrt{2} & 0 \\ -1 & -1 & 0 & \sqrt{2} \\ 1 & 1 & 0 & \sqrt{2} \end{bmatrix}$

Exercise 8.2.6 Consider $A = \begin{bmatrix} 0 & a & 0 \\ a & 0 & c \\ 0 & c & 0 \end{bmatrix}$ 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.

$P = \frac{1}{\sqrt{2}k} \begin{bmatrix} c\sqrt{2} & a & a \\ 0 & k & -k \\ -a\sqrt{2} & c & c \end{bmatrix}$

Exercise 8.2.7 Consider $A = \begin{bmatrix} 0 & 0 & a \\ 0 & b & 0 \\ a & 0 & 0 \end{bmatrix}$. 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 $A = \begin{bmatrix} b & a \\ a & b \end{bmatrix}$, 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 $A = \begin{bmatrix} b & 0 & a \\ 0 & b & 0 \\ a & 0 & b \end{bmatrix}$. 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) $q = x_1^2 + 4x_1x_2 - 2x_2^2$

---

b. $y_1 = \frac{1}{\sqrt{5}}(-x_1 + 2x_2)$ and $y_2 = \frac{1}{\sqrt{5}}(2x_1 + x_2)$; $q = -3y_1^2 + 2y_2^2$.

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.]

---

c. $\Rightarrow$ a. By Theorem 8.2.1 let $P^{-1}AP = D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ where the λ_i are the eigenvalues of A. By c. we have $\lambda_i = \pm 1$ for each i, whence $D^2 = I$. But then $A^2 = (PDP^{-1})^2 = PD^2P^{-1} = I$. Since A is symmetric this is $AA^T = I$, proving a.

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.

---

b. If $B = P^T AP = P^{-1}$, then $B^2 = P^T APP^T AP = P^T A^2 P$.

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.

If $\mathbf{x}$ and $\mathbf{y}$ are respectively columns i and j of I_n, then $\mathbf{x}^T A^T \mathbf{y} = \mathbf{x}^T A\mathbf{y}$ shows that the (i, j)-entries of A^T and A are equal.

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

---

b. $\det \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right] = 1$ and $\det \left[\begin{array}{cc} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{array} \right] = -1$ [*Remark*: These are the *only* 2×2 examples.]

d. Use the fact that $P^{-1} = P^T$ to show that $P^T(I - P) = -(I - P)^T$. Now take determinants and use the hypothesis that $\det P \neq (-1)^n$.

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 $U^T U = I$ (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 $\frac{a_{ji}}{\|\mathbf{r}_j\|^2}$.

We have $AA^T = D$, where D is diagonal with main diagonal entries $\|R_1\|^2, \dots, \|R_n\|^2$. Hence $A^{-1} = A^T D^{-1}$, and the result follows because D^{-1} has diagonal entries $1/\|R_1\|^2, \dots, 1/\|R_n\|^2$.

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.]

---

b. Because $I-A$ and $I+A$ commute, $PP^T = (I-A)(I+A)^{-1}[(I+A)^{-1}]^T(I-A)^T = (I-A)(I+A)^{-1}(I-A)^{-1}(I+A) = I$.

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 $\left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right]$ or $\left[\begin{array}{cc} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{array} \right]$ 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 **??**). We will encounter them again in Chapter **??** 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 $P^T A P = D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 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 $\lambda_j = \mathbf{x}^T A \mathbf{x} > 0$. □

Note that Theorem 8.3.2 shows that the positive definite matrices are exactly the symmetric matrices A for which the quadratic form $q = \mathbf{x}^T A \mathbf{x}$ takes only positive values.

Example 8.3.1

If U is any invertible $n \times n$ matrix, show that $A = U^T U$ is positive definite.

Solution. If $\mathbf{x}$ is in $\mathbb{R}^n$ and $\mathbf{x} \neq \mathbf{0}$, then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T (U^T U)\mathbf{x} = (U\mathbf{x})^T (U\mathbf{x}) = \|U\mathbf{x}\|^2 > 0$$

because $U\mathbf{x} \neq \mathbf{0}$ (U is invertible). Hence Theorem 8.3.2 applies.

It is remarkable that the converse to Example 8.3.1 is also true. In fact every positive definite matrix A can be factored as $A = U^T U$ where U is an upper triangular matrix with positive elements on the main diagonal. However, before verifying this, we introduce another concept that is central to any discussion of positive definite matrices.

If A is any $n \times n$ matrix, let ${}^{(r)}A$ denote the $r \times r$ submatrix in the upper left corner of A; that is, ${}^{(r)}A$ is the matrix obtained from A by deleting the last $n - r$ rows and columns. The matrices ${}^{(1)}A$, ${}^{(2)}A$, ${}^{(3)}A$, $\dots$, ${}^{(n)}A = A$ are called the **principal submatrices** of A.

Example 8.3.2

If $A = \left[\begin{array}{rrr} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{array} \right]$ then ${}^{(1)}A = [10]$, ${}^{(2)}A = \left[\begin{array}{rr} 10 & 5 \\ 5 & 3 \end{array} \right]$ and ${}^{(3)}A = A$.

Lemma 8.3.1

If A is positive definite, so is each principal submatrix ${}^{(r)}A$ for $r = 1, 2, \dots, n$.

Proof. Write $A = \left[\begin{array}{cc} {}^{(r)}A & P \\ Q & R \end{array} \right]$ in block form. If $\mathbf{y} \neq \mathbf{0}$ in $\mathbb{R}^r$, write $\mathbf{x} = \left[\begin{array}{c} \mathbf{y} \\ \mathbf{0} \end{array} \right]$ in $\mathbb{R}^n$.

Then $\mathbf{x} \neq \mathbf{0}$, so the fact that A is positive definite gives

$$0 < \mathbf{x}^T A\mathbf{x} = \left[\begin{array}{cc} \mathbf{y}^T & \mathbf{0} \end{array} \right] \left[\begin{array}{cc} {}^{(r)}A & P \\ Q & R \end{array} \right] \left[\begin{array}{c} \mathbf{y} \\ \mathbf{0} \end{array} \right] = \mathbf{y}^T ({}^{(r)}A)\mathbf{y}$$

This shows that ${}^{(r)}A$ is positive definite by Theorem 8.3.2.[5] □

If A is positive definite, Lemma 8.3.1 and Theorem 8.3.1 show that $\det({}^{(r)}A) > 0$ for every r. This proves part of the following theorem which contains the converse to Example 8.3.1, and characterizes the positive definite matrices among the symmetric ones.

[5] A similar argument shows that, if B is any matrix obtained from a positive definite matrix A by deleting certain rows and deleting the *same* columns, then B is also positive definite.

Theorem 8.3.3

The following conditions are equivalent for a symmetric $n \times n$ matrix A:

1. *A is positive definite.*
2. *$\det\left({}^{(r)}A\right) > 0$ for each $r = 1, 2, \dots, n$.*
3. *$A = U^T U$ where U is an upper triangular matrix with positive entries on the main diagonal.*

Furthermore, the factorization in (3) is unique (called the ***Cholesky factorization***[6] *of A).*

Proof. First, (3) $\Rightarrow$ (1) by Example 8.3.1, and (1) $\Rightarrow$ (2) by Lemma 8.3.1 and Theorem 8.3.1.

(2) $\Rightarrow$ (3). Assume (2) and proceed by induction on n. If $n = 1$, then $A = [a]$ where $a > 0$ by (2), so take $U = [\sqrt{a}]$. If $n > 1$, write $B = {}^{(n-1)}A$. Then B is symmetric and satisfies (2) so, by induction, we have $B = U^T U$ as in (3) where U is of size $(n-1) \times (n-1)$. Then, as A is symmetric, it has block form $A = \left[\begin{array}{cc} B & \mathbf{p} \\ \mathbf{p}^T & b \end{array}\right]$ where $\mathbf{p}$ is a column in $\mathbb{R}^{n-1}$ and b is in $\mathbb{R}$. If we write $\mathbf{x} = (U^T)^{-1}\mathbf{p}$ and $c = b - \mathbf{x}^T\mathbf{x}$, block multiplication gives

$$A = \left[\begin{array}{cc} U^T U & \mathbf{p} \\ \mathbf{p}^T & b \end{array}\right] = \left[\begin{array}{cc} U^T & 0 \\ \mathbf{x}^T & 1 \end{array}\right]\left[\begin{array}{cc} U & \mathbf{x} \\ 0 & c \end{array}\right]$$

as the reader can verify. Taking determinants and applying Theorem 3.1.5 gives $\det A = \det(U^T)\det U \cdot c = c(\det U)^2$. Hence $c > 0$ because $\det A > 0$ by (2), so the above factorization can be written

$$A = \left[\begin{array}{cc} U^T & 0 \\ \mathbf{x}^T & \sqrt{c} \end{array}\right]\left[\begin{array}{cc} U & \mathbf{x} \\ 0 & \sqrt{c} \end{array}\right]$$

Since U has positive diagonal entries, this proves (3).

As to the uniqueness, suppose that $A = U^T U = U_1^T U_1$ are two Cholesky factorizations. Now write $D = UU_1^{-1} = (U^T)^{-1}U_1^T$. Then D is upper triangular, because $D = UU_1^{-1}$, and lower triangular, because $D = (U^T)^{-1}U_1^T$, and so it is a diagonal matrix. Thus $U = DU_1$ and $U_1 = DU$, so it suffices to show that $D = I$. But eliminating U_1 gives $U = D^2 U$, so $D^2 = I$ because U is invertible. Since the diagonal entries of D are positive (this is true of U and U_1), it follows that $D = I$. □

The remarkable thing is that the matrix U in the Cholesky factorization is easy to obtain from A using row operations. The key is that Step 1 of the following algorithm is *possible* for any positive definite matrix A. A proof of the algorithm is given following Example 8.3.3.

Algorithm for the Cholesky Factorization

If A is a positive definite matrix, the Cholesky factorization $A = U^T U$ can be obtained as follows:

Step 1. Carry A to an upper triangular matrix U_1 with positive diagonal entries using row

[6] Andre-Louis Cholesky (1875–1918), was a French mathematician who died in World War I. His factorization was published in 1924 by a fellow officer.

operations each of which adds a multiple of a row to a lower row.

Step 2. Obtain U from U_1 by dividing each row of U_1 by the square root of the diagonal entry in that row.

Example 8.3.3

Find the Cholesky factorization of $A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$.

Solution. The matrix A is positive definite by Theorem 8.3.3 because $\det{}^{(1)}A = 10 > 0$, $\det{}^{(2)}A = 5 > 0$, and $\det{}^{(3)}A = \det A = 3 > 0$. Hence Step 1 of the algorithm is carried out as follows:

$$A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix} \to \begin{bmatrix} 10 & 5 & 2 \\ 0 & \frac{1}{2} & 1 \\ 0 & 1 & \frac{13}{5} \end{bmatrix} \to \begin{bmatrix} 10 & 5 & 2 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{3}{5} \end{bmatrix} = U_1$$

Now carry out Step 2 on U_1 to obtain $U = \begin{bmatrix} \sqrt{10} & \frac{5}{\sqrt{10}} & \frac{2}{\sqrt{10}} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & \frac{\sqrt{3}}{\sqrt{5}} \end{bmatrix}$.

The reader can verify that $U^T U = A$.

Proof of the Cholesky Algorithm. If A is positive definite, let $A = U^T U$ be the Cholesky factorization, and let $D = \operatorname{diag}(d_1, \dots, d_n)$ be the common diagonal of U and U^T. Then $U^T D^{-1}$ is lower triangular with ones on the diagonal (call such matrices LT-1). Hence $L = (U^T D^{-1})^{-1}$ is also LT-1, and so $I_n \to L$ by a sequence of row operations each of which adds a multiple of a row to a lower row (verify; modify columns right to left). But then $A \to LA$ by the same sequence of row operations (see the discussion preceding Theorem 2.5.1). Since $LA = [D(U^T)^{-1}][U^T U] = DU$ is upper triangular with positive entries on the diagonal, this shows that Step 1 of the algorithm is possible.

Turning to Step 2, let $A \to U_1$ as in Step 1 so that $U_1 = L_1 A$ where L_1 is LT-1. Since A is symmetric, we get

$$L_1 U_1^T = L_1 (L_1 A)^T = L_1 A^T L_1^T = L_1 A L_1^T = U_1 L_1^T \tag{8.4}$$

Let $D_1 = \operatorname{diag}(e_1, \dots, e_n)$ denote the diagonal of U_1. Then (8.4) gives $L_1(U_1^T D_1^{-1}) = U_1 L_1^T D_1^{-1}$. This is both upper triangular (right side) and LT-1 (left side), and so must equal I_n. In particular, $U_1^T D_1^{-1} = L_1^{-1}$. Now let $D_2 = \operatorname{diag}(\sqrt{e_1}, \dots, \sqrt{e_n})$, so that $D_2^2 = D_1$. If we write $U = D_2^{-1} U_1$ we have

$$U^T U = (U_1^T D_2^{-1})(D_2^{-1} U_1) = U_1^T (D_2^2)^{-1} U_1 = (U_1^T D_1^{-1}) U_1 = (L_1^{-1}) U_1 = A$$

This proves Step 2 because $U = D_2^{-1} U_1$ is formed by dividing each row of U_1 by the square root of its diagonal entry (verify). □

Exercises for 8.3

Exercise 8.3.1 Find the Cholesky decomposition of each of the following matrices.

a) $\begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix}$ b) $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 12 & 4 & 3 \\ 4 & 2 & -1 \\ 3 & -1 & 7 \end{bmatrix}$ d) $\begin{bmatrix} 20 & 4 & 5 \\ 4 & 2 & 3 \\ 5 & 3 & 5 \end{bmatrix}$

---

b. $U = \frac{\sqrt{2}}{2} \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$

d. $U = \frac{1}{30} \begin{bmatrix} 60\sqrt{5} & 12\sqrt{5} & 15\sqrt{5} \\ 0 & 6\sqrt{30} & 10\sqrt{30} \\ 0 & 0 & 5\sqrt{15} \end{bmatrix}$

Exercise 8.3.2

a. If A is positive definite, show that A^k is positive definite for all $k \geq 1$.

b. Prove the converse to (a) when k is odd.

c. Find a symmetric matrix A such that A^2 is positive definite but A is not.

---

b. If $\lambda^k > 0$, k odd, then $\lambda > 0$.

Exercise 8.3.3 Let $A = \begin{bmatrix} 1 & a \\ a & b \end{bmatrix}$. If $a^2 < b$, show that A is positive definite and find the Cholesky factorization.

Exercise 8.3.4 If A and B are positive definite and $r > 0$, show that $A + B$ and rA are both positive definite.

If $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{x}^T A\mathbf{x} > 0$ and $\mathbf{x}^T B\mathbf{x} > 0$. Hence $\mathbf{x}^T (A + B)\mathbf{x} = \mathbf{x}^T A\mathbf{x} + \mathbf{x}^T B\mathbf{x} > 0$ and $\mathbf{x}^T (rA)\mathbf{x} = r(\mathbf{x}^T A\mathbf{x}) > 0$, as $r > 0$.

Exercise 8.3.5 If A and B are positive definite, show that $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is positive definite.

Exercise 8.3.6 If A is an $n \times n$ positive definite matrix and U is an $n \times m$ matrix of rank m, show that $U^T AU$ is positive definite.

Let $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$. Then $\mathbf{x}^T (U^T AU)\mathbf{x} = (U\mathbf{x})^T A(U\mathbf{x}) > 0$ provided $U\mathbf{x} \neq 0$. But if $U = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$ and $\mathbf{x} = (x_1, x_2, \dots, x_n)$, then $U\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n \neq \mathbf{0}$ because $\mathbf{x} \neq \mathbf{0}$ and the $\mathbf{c}_i$ are independent.

Exercise 8.3.7 If A is positive definite, show that each diagonal entry is positive.

Exercise 8.3.8 Let A_0 be formed from A by deleting rows 2 and 4 and deleting columns 2 and 4. If A is positive definite, show that A_0 is positive definite.

Exercise 8.3.9 If A is positive definite, show that $A = CC^T$ where C has orthogonal columns.

Exercise 8.3.10 If A is positive definite, show that $A = C^2$ where C is positive definite.

Let $P^T AP = D = \text{diag}(\lambda_1, \dots, \lambda_n)$ where $P^T = P$. Since A is positive definite, each eigenvalue $\lambda_i > 0$. If $B = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ then $B^2 = D$, so $A = PB^2P^T = (PBP^T)^2$. Take $C = PBP^T$. Since C has eigenvalues $\sqrt{\lambda_i} > 0$, it is positive definite.

Exercise 8.3.11 Let A be a positive definite matrix. If a is a real number, show that aA is positive definite if and only if $a > 0$.

Exercise 8.3.12

a. Suppose an invertible matrix A can be factored in $\mathbf{M}_{nn}$ as $A = LDU$ where L is lower triangular with 1s on the diagonal, U is upper triangular with 1s on the diagonal, and D is diagonal with positive diagonal entries. Show that the factorization is unique: If $A = L_1D_1U_1$ is another such factorization, show that $L_1 = L$, $D_1 = D$, and $U_1 = U$.

b. Show that a matrix A is positive definite if and only if A is symmetric and admits a factorization $A = LDU$ as in (a).

---

b. If A is positive definite, use Theorem 8.3.1 to write $A = U^T U$ where U is upper triangular with positive diagonal D. Then $A = (D^{-1}U)^T D^2 (D^{-1}U)$ so $A = L_1 D_1 U_1$ is such a factorization if $U_1 = D^{-1}U$, $D_1 = D^2$, and $L_1 = U_1^T$. Conversely, let $A^T = A = LDU$ be such a factorization. Then $U^T D^T L^T = A^T = A = LDU$, so $L = U^T$ by **(a)**. Hence $A = LDL^T = V^T V$ where $V = LD_0$ and D_0 is diagonal with $D_0^2 = D$ (the matrix D_0 exists because D has positive diagonal entries). Hence A is symmetric, and it is positive definite by Example 8.3.1.

Exercise 8.3.13 Let A be positive definite and write $d_r = \det{}^{(r)}A$ for each $r = 1, 2, \ldots, n$. If U is the upper triangular matrix obtained in step 1 of the algorithm, show that the diagonal elements $u_{11}, u_{22}, \ldots, u_{nn}$ of U are given by $u_{11} = d_1$, $u_{jj} = d_j / d_{j-1}$ if $j > 1$. [*Hint*: If $LA = U$ where L is lower triangular with 1s on the diagonal, use block multiplication to show that $\det{}^{(r)}A = \det{}^{(r)}U$ for each r.]

8.4 QR-Factorization[7]

One of the main virtues of orthogonal matrices is that they can be easily inverted—the transpose is the inverse. This fact, combined with the factorization theorem in this section, provides a useful way to simplify many matrix calculations (for example, in least squares approximation).

Definition 8.6 QR-factorization

Let A be an $m \times n$ matrix with independent columns. A ***QR-factorization*** *of A expresses it as $A = QR$ where Q is $m \times n$ with orthonormal columns and R is an invertible and upper triangular matrix with positive diagonal entries.*

The importance of the factorization lies in the fact that there are computer algorithms that accomplish it with good control over round-off error, making it particularly useful in matrix calculations. The factorization is a matrix version of the Gram-Schmidt process.

Suppose $A = \left[\begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{array}\right]$ is an $m \times n$ matrix with linearly independent columns $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$. The Gram-Schmidt algorithm can be applied to these columns to provide orthogonal columns $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$ where $\mathbf{f}_1 = \mathbf{c}_1$ and

$$\mathbf{f}_k = \mathbf{c}_k - \frac{\mathbf{c}_k \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{c}_k \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\mathbf{c}_k \cdot \mathbf{f}_{k-1}}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1}$$

for each $k = 2, 3, \dots, n$. Now write $\mathbf{q}_k = \frac{1}{\|\mathbf{f}_k\|}\mathbf{f}_k$ for each k. Then $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ are orthonormal columns, and the above equation becomes

$$\|\mathbf{f}_k\|\mathbf{q}_k = \mathbf{c}_k - (\mathbf{c}_k \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{c}_k \cdot \mathbf{q}_2)\mathbf{q}_2 - \cdots - (\mathbf{c}_k \cdot \mathbf{q}_{k-1})\mathbf{q}_{k-1}$$

Using these equations, express each $\mathbf{c}_k$ as a linear combination of the $\mathbf{q}_i$:

$$\begin{array}{lcl} \mathbf{c}_1 & = & \|\mathbf{f}_1\|\mathbf{q}_1 \\ \mathbf{c}_2 & = & (\mathbf{c}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + \|\mathbf{f}_2\|\mathbf{q}_2 \\ \mathbf{c}_3 & = & (\mathbf{c}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + \|\mathbf{f}_3\|\mathbf{q}_3 \\ \vdots & & \vdots \\ \mathbf{c}_n & = & (\mathbf{c}_n \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_n \cdot \mathbf{q}_2)\mathbf{q}_2 + (\mathbf{c}_n \cdot \mathbf{q}_3)\mathbf{q}_3 + \cdots + \|\mathbf{f}_n\|\mathbf{q}_n \end{array}$$

These equations have a matrix form that gives the required factorization:

$$\begin{aligned} A &= \left[\begin{array}{ccccc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_n \end{array}\right] \\ &= \left[\begin{array}{ccccc} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \end{array}\right] \left[\begin{array}{ccccc} \|\mathbf{f}_1\| & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_1 \\ 0 & \|\mathbf{f}_2\| & \mathbf{c}_3 \cdot \mathbf{q}_2 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_2 \\ 0 & 0 & \|\mathbf{f}_3\| & \cdots & \mathbf{c}_n \cdot \mathbf{q}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \|\mathbf{f}_n\| \end{array}\right] \end{aligned} \tag{8.5}$$

Here the first factor $Q = \left[\begin{array}{ccccc} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \end{array}\right]$ has orthonormal columns, and the second factor is an $n \times n$ upper triangular matrix R with positive diagonal entries (and so is invertible). We record this in the following theorem.

[7]This section is not used elsewhere in the book

Theorem 8.4.1: QR-Factorization

Every $m \times n$ matrix A with linearly independent columns has a QR-factorization $A = QR$ where Q has orthonormal columns and R is upper triangular with positive diagonal entries.

The matrices Q and R in Theorem 8.4.1 are uniquely determined by A; we return to this below.

Example 8.4.1

Find the QR-factorization of $A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution. Denote the columns of A as $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, and observe that $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is independent. If we apply the Gram-Schmidt algorithm to these columns, the result is:

$$\mathbf{f}_1 = \mathbf{c}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{f}_2 = \mathbf{c}_2 - \tfrac{1}{2}\mathbf{f}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{f}_3 = \mathbf{c}_3 + \tfrac{1}{2}\mathbf{f}_1 - \mathbf{f}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Write $\mathbf{q}_j = \frac{1}{\|\mathbf{f}_j\|}\mathbf{f}_j$ for each j, so $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is orthonormal. Then equation (8.5) preceding Theorem 8.4.1 gives $A = QR$ where

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 \\ 0 & \frac{2}{\sqrt{6}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & 1 & 0 \\ -\sqrt{3} & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{6} \end{bmatrix}$$

$$R = \begin{bmatrix} \|\mathbf{f}_1\| & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 \\ 0 & \|\mathbf{f}_2\| & \mathbf{c}_3 \cdot \mathbf{q}_2 \\ 0 & 0 & \|\mathbf{f}_3\| \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & \frac{\sqrt{3}}{\sqrt{2}} \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 & -1 \\ 0 & \sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

The reader can verify that indeed $A = QR$.

If a matrix A has independent rows and we apply QR-factorization to A^T, the result is:

Corollary 8.4.1

If A has independent rows, then A factors uniquely as $A = LP$ where P has orthonormal rows and L is an invertible lower triangular matrix with positive main diagonal entries.

Since a square matrix with orthonormal columns is orthogonal, we have

Theorem 8.4.2

Every square, invertible matrix A has factorizations $A = QR$ and $A = LP$ where Q and P are orthogonal, R is upper triangular with positive diagonal entries, and L is lower triangular with positive diagonal entries.

Remark

In Section **??** we found how to find a best approximation $\mathbf{z}$ to a solution of a (possibly inconsistent) system $A\mathbf{x} = \mathbf{b}$ of linear equations: take $\mathbf{z}$ to be any solution of the "normal" equations $(A^TA)\mathbf{z} = A^T\mathbf{b}$. If A has independent columns this $\mathbf{z}$ is unique (A^TA is invertible by Theorem 5.4.3), so it is often desirable to compute $(A^TA)^{-1}$. This is particularly useful in least squares approximation (Section **??**). This is simplified if we have a QR-factorization of A (and is one of the main reasons for the importance of Theorem 8.4.1). For if $A = QR$ is such a factorization, then $Q^TQ = I_n$ because Q has orthonormal columns (verify), so we obtain

$$A^TA = R^TQ^TQR = R^TR$$

Hence computing $(A^TA)^{-1}$ amounts to finding R^{-1}, and this is a routine matter because R is upper triangular. Thus the difficulty in computing $(A^TA)^{-1}$ lies in obtaining the QR-factorization of A.

We conclude by proving the uniqueness of the QR-factorization.

Theorem 8.4.3

Let A be an $m \times n$ matrix with independent columns. If $A = QR$ and $A = Q_1R_1$ are QR-factorizations of A, then $Q_1 = Q$ and $R_1 = R$.

Proof. Write $Q = \left[\begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{array} \right]$ and $Q_1 = \left[\begin{array}{cccc} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_n \end{array} \right]$ in terms of their columns, and observe first that $Q^TQ = I_n = Q_1^TQ_1$ because Q and Q_1 have orthonormal columns. Hence it suffices to show that $Q_1 = Q$ (then $R_1 = Q_1^TA = Q^TA = R$). Since $Q_1^TQ_1 = I_n$, the equation $QR = Q_1R_1$ gives $Q_1^TQ = R_1R^{-1}$; for convenience we write this matrix as

$$Q_1^TQ = R_1R^{-1} = \left[\begin{array}{c} t_{ij} \end{array} \right]$$

This matrix is upper triangular with positive diagonal elements (since this is true for R and R_1), so $t_{ii} > 0$ for each i and $t_{ij} = 0$ if $i > j$. On the other hand, the $(i,\ j)$-entry of Q_1^TQ is $\mathbf{d}_i^T\mathbf{c}_j = \mathbf{d}_i \cdot \mathbf{c}_j$, so we have $\mathbf{d}_i \cdot \mathbf{c}_j = t_{ij}$ for all i and j. But each $\mathbf{c}_j$ is in $\text{span}\,\{\mathbf{d}_1,\ \mathbf{d}_2,\ \dots,\ \mathbf{d}_n\}$ because $Q = Q_1(R_1R^{-1})$. Hence the expansion theorem gives

$$\mathbf{c}_j = (\mathbf{d}_1 \cdot \mathbf{c}_j)\mathbf{d}_1 + (\mathbf{d}_2 \cdot \mathbf{c}_j)\mathbf{d}_2 + \cdots + (\mathbf{d}_n \cdot \mathbf{c}_j)\mathbf{d}_n = t_{1j}\mathbf{d}_1 + t_{2j}\mathbf{d}_2 + \cdots + t_{jj}\mathbf{d}_i$$

because $\mathbf{d}_i \cdot \mathbf{c}_j = t_{ij} = 0$ if $i > j$. The first few equations here are

$$\begin{array}{lcl}
\mathbf{c}_1 & = & t_{11}\mathbf{d}_1 \\
\mathbf{c}_2 & = & t_{12}\mathbf{d}_1 + t_{22}\mathbf{d}_2 \\
\mathbf{c}_3 & = & t_{13}\mathbf{d}_1 + t_{23}\mathbf{d}_2 + t_{33}\mathbf{d}_3 \\
\mathbf{c}_4 & = & t_{14}\mathbf{d}_1 + t_{24}\mathbf{d}_2 + t_{34}\mathbf{d}_3 + t_{44}\mathbf{d}_4 \\
\vdots & & \vdots
\end{array}$$

The first of these equations gives $1 = \|\mathbf{c}_1\| = \|t_{11}\mathbf{d}_1\| = |t_{11}|\|\mathbf{d}_1\| = t_{11}$, whence $\mathbf{c}_1 = \mathbf{d}_1$. But then we have $t_{12} = \mathbf{d}_1 \cdot \mathbf{c}_2 = \mathbf{c}_1 \cdot \mathbf{c}_2 = 0$, so the second equation becomes $\mathbf{c}_2 = t_{22}\mathbf{d}_2$. Now a similar argument gives $\mathbf{c}_2 = \mathbf{d}_2$, and then $t_{13} = 0$ and $t_{23} = 0$ follows in the same way. Hence $\mathbf{c}_3 = t_{33}\mathbf{d}_3$ and $\mathbf{c}_3 = \mathbf{d}_3$. Continue in this way to get $\mathbf{c}_i = \mathbf{d}_i$ for all i. This means that $Q_1 = Q$, which is what we wanted. □

Exercises for 8.4

Exercise 8.4.1 In each case find the QR-factorization of A.

a) $A = \left[\begin{array}{rr} 1 & -1 \\ -1 & 0 \end{array}\right]$ b) $A = \left[\begin{array}{rr} 2 & 1 \\ 1 & 1 \end{array}\right]$

c) $A = \left[\begin{array}{rrr} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$ d) $A = \left[\begin{array}{rrr} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{array}\right]$

---

b. $Q = \frac{1}{\sqrt{5}}\left[\begin{array}{rr} 2 & -1 \\ 1 & 2 \end{array}\right]$, $R = \frac{1}{\sqrt{5}}\left[\begin{array}{rr} 5 & 3 \\ 0 & 1 \end{array}\right]$

d. $Q = \frac{1}{\sqrt{3}}\left[\begin{array}{rrr} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{array}\right]$,

$R = \frac{1}{\sqrt{3}}\left[\begin{array}{rrr} 3 & 0 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{array}\right]$

Exercise 8.4.2 Let A and B denote matrices.

a. If A and B have independent columns, show that AB has independent columns. [*Hint*: Theorem 5.4.3.]

b. Show that A has a QR-factorization if and only if A has independent columns.

c. If AB has a QR-factorization, show that the same is true of B but not necessarily A. [*Hint*: Consider AA^T where $A = \left[\begin{array}{rrr} 1 & 0 & 0 \\ 1 & 1 & 1 \end{array}\right]$.]

If A has a QR-factorization, use (a). For the converse use Theorem 8.4.1.

Exercise 8.4.3 If R is upper triangular and invertible, show that there exists a diagonal matrix D with diagonal entries ± 1 such that $R_1 = DR$ is invertible, upper triangular, and has positive diagonal entries.

Exercise 8.4.4 If A has independent columns, let $A = QR$ where Q has orthonormal columns and R is invertible and upper triangular. [Some authors call *this* a QR-factorization of A.] Show that there is a diagonal matrix D with diagonal entries ± 1 such that $A = (QD)(DR)$ is the QR-factorization of A. [*Hint*: Preceding exercise.]

8.5 Computing Eigenvalues

In practice, the problem of finding eigenvalues of a matrix is virtually never solved by finding the roots of the characteristic polynomial. This is difficult for large matrices and iterative methods are much better. Two such methods are described briefly in this section.

The Power Method

In Chapter 3 our initial rationale for diagonalizing matrices was to be able to compute the powers of a square matrix, and the eigenvalues were needed to do this. In this section, we are interested in efficiently computing eigenvalues, and it may come as no surprise that the first method we discuss uses the powers of a matrix.

Recall that an eigenvalue λ of an $n \times n$ matrix A is called a **dominant eigenvalue** if λ has multiplicity 1, and

$$|\lambda| > |\mu| \quad \text{for all eigenvalues } \mu \neq \lambda$$

Any corresponding eigenvector is called a **dominant eigenvector** of A. When such an eigenvalue exists, one technique for finding it is as follows: Let $\mathbf{x}_0$ in $\mathbb{R}^n$ be a first approximation to a dominant eigenvector λ, and compute successive approximations $\mathbf{x}_1, \mathbf{x}_2, \dots$ as follows:

$$\mathbf{x}_1 = A\mathbf{x}_0 \quad \mathbf{x}_2 = A\mathbf{x}_1 \quad \mathbf{x}_3 = A\mathbf{x}_2 \quad \cdots$$

In general, we define

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for each } k \geq 0$$

If the first estimate $\mathbf{x}_0$ is good enough, these vectors $\mathbf{x}_n$ will approximate the dominant eigenvector λ (see below). This technique is called the **power method** (because $\mathbf{x}_k = A^k\mathbf{x}_0$ for each $k \geq 1$). Observe that if $\mathbf{z}$ is any eigenvector corresponding to λ, then

$$\frac{\mathbf{z}\cdot(A\mathbf{z})}{\|\mathbf{z}\|^2} = \frac{\mathbf{z}\cdot(\lambda\mathbf{z})}{\|\mathbf{z}\|^2} = \lambda$$

Because the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ approximate dominant eigenvectors, this suggests that we define the **Rayleigh quotients** as follows:

$$r_k = \frac{\mathbf{x}_k\cdot\mathbf{x}_{k+1}}{\|\mathbf{x}_k\|^2} \quad \text{for } k \geq 1$$

Then the numbers r_k approximate the dominant eigenvalue λ.

Example 8.5.1

Use the power method to approximate a dominant eigenvector and eigenvalue of $A = \left[\begin{array}{rr} 1 & 1 \\ 2 & 0 \end{array} \right]$.

Solution. The eigenvalues of A are 2 and -1, with eigenvectors $\left[\begin{array}{r} 1 \\ 1 \end{array} \right]$ and $\left[\begin{array}{r} 1 \\ -2 \end{array} \right]$. Take

$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ as the first approximation and compute $\mathbf{x}_1, \mathbf{x}_2, \dots,$ successively, from $\mathbf{x}_1 = A\mathbf{x}_0, \mathbf{x}_2 = A\mathbf{x}_1, \dots$. The result is

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 11 \\ 10 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 21 \\ 22 \end{bmatrix}, \quad \dots$$

These vectors are approaching scalar multiples of the dominant eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Moreover, the Rayleigh quotients are

$$r_1 = \tfrac{7}{5},\ r_2 = \tfrac{27}{13},\ r_3 = \tfrac{115}{61},\ r_4 = \tfrac{451}{221},\ \dots$$

and these are approaching the dominant eigenvalue 2.

To see why the power method works, let $\lambda_1, \lambda_2, \dots, \lambda_m$ be eigenvalues of A with λ_1 dominant and let $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ be corresponding eigenvectors. What is required is that the first approximation $\mathbf{x}_0$ be a linear combination of these eigenvectors:

$$\mathbf{x}_0 = a_1\mathbf{y}_1 + a_2\mathbf{y}_2 + \cdots + a_m\mathbf{y}_m \quad \text{with } a_1 \neq 0$$

If $k \geq 1$, the fact that $\mathbf{x}_k = A^k\mathbf{x}_0$ and $A^k\mathbf{y}_i = \lambda_i^k\mathbf{y}_i$ for each i gives

$$\mathbf{x}_k = a_1\lambda_1^k\mathbf{y}_1 + a_2\lambda_2^k\mathbf{y}_2 + \cdots + a_m\lambda_m^k\mathbf{y}_m \quad \text{for } k \geq 1$$

Hence

$$\tfrac{1}{\lambda_1^k}\mathbf{x}_k = a_1\mathbf{y}_1 + a_2\left(\tfrac{\lambda_2}{\lambda_1}\right)^k\mathbf{y}_2 + \cdots + a_m\left(\tfrac{\lambda_m}{\lambda_1}\right)^k\mathbf{y}_m$$

The right side approaches $a_1\mathbf{y}_1$ as k increases because λ_1 is dominant $\left(\left|\frac{\lambda_i}{\lambda_1}\right| < 1 \text{ for each } i > 1\right)$. Because $a_1 \neq 0$, this means that $\mathbf{x}_k$ approximates the dominant eigenvector $a_1\lambda_1^k\mathbf{y}_1$.

The power method requires that the first approximation $\mathbf{x}_0$ be a linear combination of eigenvectors. (In Example 8.5.1 the eigenvectors form a basis of $\mathbb{R}^2$.) But even in this case the method fails if $a_1 = 0$, where a_1 is the coefficient of the dominant eigenvector (try $\mathbf{x}_0 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ in Example 8.5.1). In general, the rate of convergence is quite slow if any of the ratios $\left|\frac{\lambda_i}{\lambda_1}\right|$ is near 1. Also, because the method requires repeated multiplications by A, it is not recommended unless these multiplications are easy to carry out (for example, if most of the entries of A are zero).

QR-Algorithm

A much better method for approximating the eigenvalues of an invertible matrix A depends on the factorization (using the Gram-Schmidt algorithm) of A in the form

$$A = QR$$

where Q is orthogonal and R is invertible and upper triangular (see Theorem 8.4.2). The **QR-algorithm** uses this repeatedly to create a sequence of matrices $A_1 = A,\ A_2,\ A_3,\ \dots,$ as follows:

1. Define $A_1 = A$ and factor it as $A_1 = Q_1R_1$.
2. Define $A_2 = R_1Q_1$ and factor it as $A_2 = Q_2R_2$.
3. Define $A_3 = R_2Q_2$ and factor it as $A_3 = Q_3R_3$.

$\vdots$

In general, A_k is factored as $A_k = Q_kR_k$ and we define $A_{k+1} = R_kQ_k$. Then A_{k+1} is similar to A_k [in fact, $A_{k+1} = R_kQ_k = (Q_k^{-1}A_k)Q_k$], and hence each A_k has the same eigenvalues as A. If the eigenvalues of A are real and have distinct absolute values, the remarkable thing is that the sequence of matrices $A_1,\ A_2,\ A_3,\ \dots$ converges to an upper triangular matrix with these eigenvalues on the main diagonal. [See below for the case of complex eigenvalues.]

Example 8.5.2

If $A = \left[\begin{array}{rr} 1 & 1 \\ 2 & 0 \end{array} \right]$ as in Example 8.5.1, use the QR-algorithm to approximate the eigenvalues.

Solution. The matrices A_1, A_2, and A_3 are as follows:

$$A_1 = \left[\begin{array}{rr} 1 & 1 \\ 2 & 0 \end{array} \right] = Q_1R_1 \quad \text{where } Q_1 = \frac{1}{\sqrt{5}}\left[\begin{array}{rr} 1 & 2 \\ 2 & -1 \end{array} \right] \text{ and } R_1 = \frac{1}{\sqrt{5}}\left[\begin{array}{rr} 5 & 1 \\ 0 & 2 \end{array} \right]$$

$$A_2 = \frac{1}{5}\left[\begin{array}{rr} 7 & 9 \\ 4 & -2 \end{array} \right] = \left[\begin{array}{rr} 1.4 & -1.8 \\ -0.8 & -0.4 \end{array} \right] = Q_2R_2$$

$$\text{where } Q_2 = \frac{1}{\sqrt{65}}\left[\begin{array}{rr} 7 & 4 \\ 4 & -7 \end{array} \right] \text{ and } R_2 = \frac{1}{\sqrt{65}}\left[\begin{array}{rr} 13 & 11 \\ 0 & 10 \end{array} \right]$$

$$A_3 = \frac{1}{13}\left[\begin{array}{rr} 27 & -5 \\ 8 & -14 \end{array} \right] = \left[\begin{array}{rr} 2.08 & -0.38 \\ 0.62 & -1.08 \end{array} \right]$$

This is converging to $\left[\begin{array}{rr} 2 & * \\ 0 & -1 \end{array} \right]$ and so is approximating the eigenvalues 2 and -1 on the main diagonal.

It is beyond the scope of this book to pursue a detailed discussion of these methods. The reader is referred to J. M. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford, England: Oxford University

Press, 1965) or G. W. Stewart, *Introduction to Matrix Computations* (New York: Academic Press, 1973). We conclude with some remarks on the QR-algorithm.

Shifting. Convergence is accelerated if, at stage k of the algorithm, a number s_k is chosen and $A_k - s_k I$ is factored in the form $Q_k R_k$ rather than A_k itself. Then

$$Q_k^{-1} A_k Q_k = Q_k^{-1}(Q_k R_k + s_k I)Q_k = R_k Q_k + s_k I$$

so we take $A_{k+1} = R_k Q_k + s_k I$. If the shifts s_k are carefully chosen, convergence can be greatly improved.

Preliminary Preparation. A matrix such as

$$\left[\begin{array}{ccccc} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{array}\right]$$

is said to be in **upper Hessenberg** form, and the QR-factorizations of such matrices are greatly simplified. Given an $n \times n$ matrix A, a series of orthogonal matrices $H_1, H_2, \dots, H_m$ (called **Householder matrices**) can be easily constructed such that

$$B = H_m^T \cdots H_1^T A H_1 \cdots H_m$$

is in upper Hessenberg form. Then the QR-algorithm can be efficiently applied to B and, because B is similar to A, it produces the eigenvalues of A.

Complex Eigenvalues. If some of the eigenvalues of a real matrix A are not real, the QR-algorithm converges to a block upper triangular matrix where the diagonal blocks are either 1×1 (the real eigenvalues) or 2×2 (each providing a pair of conjugate complex eigenvalues of A).

Exercises for 8.5

Exercise 8.5.1 In each case, find the exact eigenvalues and determine corresponding eigenvectors. Then start with $\mathbf{x}_0 = \left[\begin{array}{c} 1 \\ 1 \end{array}\right]$ and compute $\mathbf{x}_4$ and r_3 using the power method.

a) $A = \left[\begin{array}{rr} 2 & -4 \\ -3 & 3 \end{array}\right]$ b) $A = \left[\begin{array}{rr} 5 & 2 \\ -3 & -2 \end{array}\right]$

c) $A = \left[\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right]$ d) $A = \left[\begin{array}{cc} 3 & 1 \\ 1 & 0 \end{array}\right]$

---

b. Eigenvalues 4, -1; eigenvectors $\left[\begin{array}{r} 2 \\ -1 \end{array}\right]$, $\left[\begin{array}{r} 1 \\ -3 \end{array}\right]$; $\mathbf{x}_4 = \left[\begin{array}{r} 409 \\ -203 \end{array}\right]$; $r_3 = 3.94$

d. Eigenvalues $\lambda_1 = \frac{1}{2}(3+\sqrt{13})$, $\lambda_2 = \frac{1}{2}(3-\sqrt{13})$; eigenvectors $\left[\begin{array}{c} \lambda_1 \\ 1 \end{array}\right]$, $\left[\begin{array}{c} \lambda_2 \\ 1 \end{array}\right]$; $\mathbf{x}_4 = \left[\begin{array}{c} 142 \\ 43 \end{array}\right]$; $r_3 = 3.3027750$ (The true value is $\lambda_1 = 3.3027756$, to seven decimal places.)

Exercise 8.5.2 In each case, find the exact eigenvalues and then approximate them using the QR-algorithm.

a) $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ b) $A = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$

---

b. Eigenvalues $\lambda_1 = \frac{1}{2}(3+\sqrt{13}) = 3.302776$, $\lambda_2 = \frac{1}{2}(3-\sqrt{13}) = -0.302776$ $A_1 = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$, $Q_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}$, $R_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} 10 & 3 \\ 0 & -1 \end{bmatrix}$

$$A_2 = \frac{1}{10}\begin{bmatrix} 33 & -1 \\ -1 & -3 \end{bmatrix},$$

$$Q_2 = \frac{1}{\sqrt{1090}}\begin{bmatrix} 33 & 1 \\ -1 & 33 \end{bmatrix},$$

$$R_2 = \frac{1}{\sqrt{1090}}\begin{bmatrix} 109 & -3 \\ 0 & -10 \end{bmatrix}$$

$$A_3 = \frac{1}{109}\begin{bmatrix} 360 & 1 \\ 1 & -33 \end{bmatrix}$$

$$= \begin{bmatrix} 3.302775 & 0.009174 \\ 0.009174 & -0.302775 \end{bmatrix}$$

Exercise 8.5.3 Apply the power method to $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, starting at $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Does it converge? Explain.

Exercise 8.5.4 If A is symmetric, show that each matrix A_k in the QR-algorithm is also symmetric. Deduce that they converge to a diagonal matrix.

Use induction on k. If $k = 1$, $A_1 = A$. In general $A_{k+1} = Q_k^{-1}A_kQ_k = Q_k^TA_kQ_k$, so the fact that $A_k^T = A_k$ implies $A_{k+1}^T = A_{k+1}$. The eigenvalues of A are all real (Theorem 5.5.5), so the A_k converge to an upper triangular matrix T. But T must also be symmetric (it is the limit of symmetric matrices), so it is diagonal.

Exercise 8.5.5 Apply the QR-algorithm to $A = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}$. Explain.

Exercise 8.5.6 Given a matrix A, let A_k, Q_k, and R_k, $k \geq 1$, be the matrices constructed in the QR-algorithm. Show that $A_k = (Q_1Q_2\cdots Q_k)(R_k\cdots R_2R_1)$ for each $k \geq 1$ and hence that this is a QR-factorization of A_k.
[*Hint*: Show that $Q_kR_k = R_{k-1}Q_{k-1}$ for each $k \geq 2$, and use this equality to compute $(Q_1Q_2\cdots Q_k)(R_k\cdots R_2R_1)$ "from the centre out." Use the fact that $(AB)^{n+1} = A(BA)^nB$ for any square matrices A and B.]

8.6 The Singular Value Decomposition

When working with a square matrix A it is clearly useful to be able to "diagonalize" A, that is to find a factorization $A = Q^{-1}DQ$ where Q is invertible and D is diagonal. Unfortunately such a factorization may not exist for A. However, even if A is not square gaussian elimination provides a factorization of the form $A = PDQ$ where P and Q are invertible and D is diagonal—the Smith Normal form (Theorem 2.5.3). However, if A is real we can choose P and Q to be *orthogonal* real matrices and D to be real. Such a factorization is called a **singular value decomposition (SVD)** for A, one of the most useful tools in applied linear algebra. In this Section we show how to explicitly compute an SVD for any real matrix A, and illustrate some of its many applications.

We need a fact about two subspaces associated with an $m \times n$ matrix A:

$$\operatorname{im} A = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\} \qquad \text{and} \qquad \operatorname{col} A = \operatorname{span}\{\mathbf{a} \mid \mathbf{a} \text{ is a column of } A\}$$

Then $\operatorname{im} A$ is called the **image** of A (so named because of the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{x} \mapsto A\mathbf{x}$); and $\operatorname{col} A$ is called the **column space** of A (Definition 5.10). Surprisingly, these spaces are equal:

Lemma 8.6.1

For any $m \times n$ matrix A, $\operatorname{im} A = \operatorname{col} A$.

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ in terms of its columns. Let $\mathbf{x} \in \operatorname{im} A$, say $\mathbf{x} = A\mathbf{y}$, $\mathbf{y}$ in $\mathbb{R}^n$. If $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$, then $A\mathbf{y} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \cdots + y_n\mathbf{a}_n \in \operatorname{col} A$ by Definition 2.5. This shows that $\operatorname{im} A \subseteq \operatorname{col} A$. For the other inclusion, each $\mathbf{a}_k = A\mathbf{e}_k$ where $\mathbf{e}_k$ is column k of I_n. □

8.6.1. Singular Value Decompositions

We know a lot about any real symmetric matrix: Its eigenvalues are real (Theorem 5.5.7), and it is orthogonally diagonalizable by the Principal Axes Theorem (Theorem 8.2.2). So for any real matrix A (square or not), the fact that both A^TA and AA^T are real and symmetric suggests that we can learn a lot about A by studying them. This section shows just how true this is.

The following Lemma reveals some similarities between A^TA and AA^T which simplify the statement and the proof of the SVD we are constructing.

Lemma 8.6.2

Let A be a real $m \times n$ matrix. Then:

1. *The eigenvalues of A^TA and AA^T are real and non-negative.*
2. *A^TA and AA^T have the same set of positive eigenvalues.*

Proof.

1. Let λ be an eigenvalue of A^TA, with eigenvector $\mathbf{0} \neq \mathbf{q} \in \mathbb{R}^n$. Then:
$$\|A\mathbf{q}\|^2 = (A\mathbf{q})^T(A\mathbf{q}) = \mathbf{q}^T(A^TA\mathbf{q}) = \mathbf{q}^T(\lambda\mathbf{q}) = \lambda(\mathbf{q}^T\mathbf{q}) = \lambda\|\mathbf{q}\|^2$$
Then (1.) follows for A^TA, and the case AA^T follows by replacing A by A^T.

2. Write $N(B)$ for the set of positive eigenvalues of a matrix B. We must show that $N(A^TA) = N(AA^T)$. If $\lambda \in N(A^TA)$ with eigenvector $\mathbf{0} \neq \mathbf{q} \in \mathbb{R}^n$, then $A\mathbf{q} \in \mathbb{R}^m$ and
$$AA^T(A\mathbf{q}) = A[(A^TA)\mathbf{q}] = A(\lambda\mathbf{q}) = \lambda(A\mathbf{q})$$
Moreover, $A\mathbf{q} \neq \mathbf{0}$ since $A^TA\mathbf{q} = \lambda\mathbf{q} \neq \mathbf{0}$ and both $\lambda \neq 0$ and $\mathbf{q} \neq \mathbf{0}$. Hence λ is an eigenvalue of AA^T, proving $N(A^TA) \subseteq N(AA^T)$. For the other inclusion replace A by A^T.

□

To analyze an $m \times n$ matrix A we have two symmetric matrices to work with: A^TA and AA^T. In view of Lemma 8.6.2, we choose A^TA (sometimes called the **Gram** matrix of A), and derive a series of facts which we will need. This narrative is a bit long, but trust that it will be worth the effort. We parse it out in several steps:

1. The $n \times n$ matrix A^TA is real and symmetric so, by the Principal Axes Theorem 8.2.2, let $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\} \subseteq \mathbb{R}^n$ be an orthonormal basis of eigenvectors of A^TA, with corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. By Lemma 8.6.2(1), λ_i is real for each i and $\lambda_i \geq 0$. By re-ordering the $\mathbf{q}_i$ we may (and do) assume that
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \quad \text{and}^{8} \quad \lambda_i = 0 \text{ if } i > r \tag{i}$$
By Theorems 8.2.1 and 3.3.4, the matrix
$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \text{ is orthogonal} \quad \text{and} \quad \text{orthogonally diagonalizes } A^TA \tag{ii}$$

2. Even though the λ_i are the eigenvalues of A^TA, the number r in (**i**) turns out to be $\text{rank}\, A$. To understand why, consider the vectors $A\mathbf{q}_i \in \text{im}\, A$. For all i, j:
$$A\mathbf{q}_i \cdot A\mathbf{q}_j = (A\mathbf{q}_i)^TA\mathbf{q}_j = \mathbf{q}_i^T(A^TA)\mathbf{q}_j = \mathbf{q}_i^T(\lambda_j\mathbf{q}_j) = \lambda_j(\mathbf{q}_i^T\mathbf{q}_j) = \lambda_j(\mathbf{q}_i \cdot \mathbf{q}_j)$$
Because $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ is an orthonormal set, this gives
$$A\mathbf{q}_i \cdot A\mathbf{q}_j = 0 \text{ if } i \neq j \qquad \text{and} \qquad \|A\mathbf{q}_i\|^2 = \lambda_i\|\mathbf{q}_i\|^2 = \lambda_i \text{ for each } i \tag{iii}$$
We can extract two conclusions from (**iii**) and (**i**):
$$\{A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r\} \subseteq \text{im}\, A \text{ is an orthogonal set} \quad \text{and} \quad A\mathbf{q}_i = \mathbf{0} \text{ if } i > r \tag{iv}$$
With this write $U = \text{span}\,\{A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r\} \subseteq \text{im}\, A$; we claim that $U = \text{im}\, A$, that is $\text{im}\, A \subseteq U$. For this we must show that $A\mathbf{x} \in U$ for each $\mathbf{x} \in \mathbb{R}^n$. Since $\{\mathbf{q}_1, \dots, \mathbf{q}_r, \dots, \mathbf{q}_n\}$ is a basis of

[8] Of course they could *all* be positive ($r = n$) or *all* zero (so $A^TA = 0$, and hence $A = 0$ by Exercise 5.3.9).

$\mathbb{R}^n$ (it is orthonormal), we can write $\mathbf{x}_k = t_1\mathbf{q}_1 + \cdots + t_r\mathbf{q}_r + \cdots + t_n\mathbf{q}_n$ where each $t_j \in \mathbb{R}$. Then, using (**iv**) we obtain

$$A\mathbf{x} = t_1A\mathbf{q}_1 + \cdots + t_rA\mathbf{q}_r + \cdots + t_nA\mathbf{q}_n = t_1A\mathbf{q}_1 + \cdots + t_rA\mathbf{q}_r \in U$$

This shows that $U = \operatorname{im} A$, and so

$$\{A\mathbf{q}_1,\ A\mathbf{q}_2,\ \dots,\ A\mathbf{q}_r\} \text{ is an } \textit{orthogonal} \text{ basis of } \operatorname{im}(A) \tag{v}$$

But $\operatorname{col} A = \operatorname{im} A$ by Lemma 8.6.1, and $\operatorname{rank} A = \dim(\operatorname{col} A)$ by Theorem 5.4.1, so

$$\operatorname{rank} A = \dim(\operatorname{col} A) = \dim(\operatorname{im} A) \overset{(\mathbf{v})}{=} r \tag{vi}$$

3. Before proceeding, some definitions are in order:

Definition 8.7

The real numbers $\sigma_i = \sqrt{\lambda_i} \overset{(iii)}{=} \|A\bar{\mathbf{q}}_i\|$ for $i = 1,\ 2,\ \dots,\ n$, are called the ***singular values*** *of the matrix A.*

Clearly $\sigma_1,\ \sigma_2,\ \dots,\ \sigma_r$ are the *positive* singular values of A. By (**i**) we have

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \qquad \text{and} \qquad \sigma_i = 0 \text{ if } i > r \tag{vii}$$

With (**vi**) this makes the following definitions depend only upon A.

Definition 8.8

Let A be a real, $m \times n$ matrix of rank r, with positive singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_i = 0$ if $i > r$. Define:

$$D_A = \operatorname{diag}(\sigma_1,\ \dots,\ \sigma_r) \qquad \textit{and} \qquad \Sigma_A = \begin{bmatrix} D_A & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$$

Here Σ_A is in block form and is called the ***singular matrix*** *of A.*

The singular values σ_i and the matrices D_A and Σ_A will be referred to frequently below.

4. Returning to our narrative, normalize the vectors $A\mathbf{q}_1,\ A\mathbf{q}_2,\ \dots,\ A\mathbf{q}_r$, by defining

$$\mathbf{p}_i = \frac{1}{\|A\mathbf{q}_i\|}A\mathbf{q}_i \in \mathbb{R}^m \quad \text{for each } i = 1,\ 2,\ \dots,\ r \tag{viii}$$

By (**v**) and Lemma 8.6.1, we conclude that

$$\{\mathbf{p}_1,\ \mathbf{p}_2,\ \dots,\ \mathbf{p}_r\} \text{ is an } \textit{orthonormal} \text{ basis of } \operatorname{col} A \subseteq \mathbb{R}^m \tag{ix}$$

Employing the Gram-Schmidt algorithm (or otherwise), construct $\mathbf{p}_{r+1},\ \dots,\ \mathbf{p}_m$ so that

$$\{\mathbf{p}_1,\ \dots,\ \mathbf{p}_r,\ \dots,\ \mathbf{p}_m\} \text{ is an orthonormal basis of } \mathbb{R}^m \tag{x}$$

5. By (**x**) and (**ii**) we have *two* orthogonal matrices

$$P = \left[\begin{array}{ccccc} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \cdots & \mathbf{p}_m \end{array}\right] \text{ of size } m \times m \quad \text{and} \quad Q = \left[\begin{array}{ccccc} \mathbf{q}_1 & \cdots & \mathbf{q}_r & \cdots & \mathbf{q}_n \end{array}\right] \text{ of size } n \times n$$

These matrices are related. In fact we have:

$$\sigma_i \mathbf{p}_i = \sqrt{\lambda_i}\mathbf{p}_i \overset{\text{(iii)}}{=} \|A\mathbf{q}_i\|\mathbf{p}_i \overset{\text{(viii)}}{=} A\mathbf{q}_i \quad \text{for each } i = 1,\ 2,\ \ldots,\ r \tag{xi}$$

This yields the following expression for AQ in terms of its columns:

$$AQ = \left[\begin{array}{cccccc} A\mathbf{q}_1 & \cdots & A\mathbf{q}_r & A\mathbf{q}_{r+1} & \cdots & A\mathbf{q}_n \end{array}\right] \overset{\text{(iv)}}{=} \left[\begin{array}{cccccc} \sigma_1\mathbf{p}_1 & \cdots & \sigma_r\mathbf{p}_r & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right] \tag{xii}$$

Then we compute:

$$\begin{aligned} P\Sigma_A &= \left[\begin{array}{cccccc} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \mathbf{p}_{r+1} & \cdots & \mathbf{p}_m \end{array}\right] \left[\begin{array}{ccccccc} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right] \\ &= \left[\begin{array}{cccccc} \sigma_1\mathbf{p}_1 & \cdots & \sigma_r\mathbf{p}_r & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right] \\ &\overset{\text{(xii)}}{=} AQ \end{aligned}$$

Finally, as $Q^{-1} = Q^T$ it follows that $A = P\Sigma_A Q^T$.

With this we can state the main theorem of this Section.

Theorem 8.6.1

Let A be a real $m \times n$ matrix, and let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the positive singular values of A. Then r is the rank of A and we have the factorization

$$A = P\Sigma_A Q^T \qquad \textit{where } P \textit{ and } Q \textit{ are orthogonal matrices}$$

The factorization $A = P\Sigma_A Q^T$ in Theorem 8.6.1, where P and Q are orthogonal matrices, is called a *Singular Value Decomposition* (*SVD*) of A. This decomposition is not unique. For example if $r < m$ then the vectors $\mathbf{p}_{r+1},\ \ldots,\ \mathbf{p}_m$ can be *any* extension of $\{\mathbf{p}_1,\ \ldots,\ \mathbf{p}_r\}$ to an orthonormal basis of $\mathbb{R}^m$, and each will lead to a different matrix P in the decomposition. For a more dramatic example, if $A = I_n$ then $\Sigma_A = I_n$, and $A = P\Sigma_A P^T$ is a SVD of A for *any* orthogonal $n \times n$ matrix P.

Example 8.6.1

Find a singular value decomposition for $A = \left[\begin{array}{rrr} 1 & 0 & 1 \\ -1 & 1 & 0 \end{array}\right]$.

Solution. We have $A^TA = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, so the characteristic polynomial is

$$c_{A^TA}(x) = \det \begin{bmatrix} x-2 & 1 & -1 \\ 1 & x-1 & 0 \\ -1 & 0 & x-1 \end{bmatrix} = (x-3)(x-1)x$$

Hence the eigenvalues of A^TA (in descending order) are $\lambda_1 = 3$, $\lambda_2 = 1$ and $\lambda_3 = 0$ with, respectively, unit eigenvectors

$$\mathbf{q}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{q}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

It follows that the orthogonal matrix Q in Theorem 8.6.1 is

$$Q = \left[\begin{array}{ccc} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{array}\right] = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & 0 & -\sqrt{2} \\ -1 & \sqrt{3} & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \end{bmatrix}$$

The singular values here are $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$ and $\sigma_3 = 0$, so $\operatorname{rank}(A) = 2$—clear in this case—and the singular matrix is

$$\Sigma_A = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

So it remains to find the 2×2 orthogonal matrix P in Theorem 8.6.1. This involves the vectors

$$A\mathbf{q}_1 = \frac{\sqrt{6}}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad A\mathbf{q}_2 = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad A\mathbf{q}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Normalize $A\mathbf{q}_1$ and $A\mathbf{q}_2$ to get

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

In this case, $\{\mathbf{p}_1, \mathbf{p}_2\}$ is *already* a basis of $\mathbb{R}^2$ (so the Gram-Schmidt algorithm is not needed), and we have the 2×2 orthogonal matrix

$$P = \left[\begin{array}{cc} \mathbf{p}_1 & \mathbf{p}_2 \end{array}\right] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Finally (by Theorem 8.6.1) the singular value decomposition for A is

$$A = P\Sigma_A Q^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & -1 & 1 \\ 0 & \sqrt{3} & \sqrt{3} \\ -\sqrt{2} & -\sqrt{2} & \sqrt{2} \end{bmatrix}$$

Of course this can be confirmed by direct matrix multiplication.

Thus, computing an SVD for a real matrix A is a routine matter, and we now describe a systematic procedure for doing so.

SVD Algorithm

Given a real $m \times n$ matrix A, find an SVD $A = P\Sigma_A Q^T$ as follows:

1. *Use the Diagonalization Algorithm (see page 188) to find the (real and non-negative) eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A^TA with corresponding (orthonormal) eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. Reorder the $\mathbf{q}_i$ (if necessary) to ensure that the nonzero eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\lambda_i = 0$ if $i > r$.*

2. *The integer r is the rank of the matrix A.*

3. *The $n \times n$ orthogonal matrix Q in the SVD is $Q = \left[\begin{array}{cccc} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{array} \right]$.*

4. *Define $\mathbf{p}_i = \frac{1}{\|A\mathbf{q}_i\|}A\mathbf{q}_i$ for $i = 1, 2, \dots, r$ (where r is as in step 1). Then $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_r\}$ is orthonormal in $\mathbb{R}^m$ so (using Gram-Schmidt or otherwise) extend it to an orthonormal basis $\{\mathbf{p}_1, \dots, \mathbf{p}_r, \dots, \mathbf{p}_m\}$ in $\mathbb{R}^m$.*

5. *The $m \times m$ orthogonal matrix P in the SVD is $P = \left[\begin{array}{ccccc} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \cdots & \mathbf{p}_m \end{array} \right]$.*

6. *The* singular values *for A are $\sigma_1, \sigma_2, \dots, \sigma_n$ where $\sigma_i = \sqrt{\lambda_i}$ for each i. Hence the nonzero singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and so the singular matrix of A in the SVD is $\Sigma_A = \left[\begin{array}{cc} \text{diag}(\sigma_1, \dots, \sigma_r) & 0 \\ 0 & 0 \end{array} \right]_{m \times n}$.*

7. *Thus $A = P\Sigma Q^T$ is a SVD for A.*

In practise the singular values σ_i, the matrices P and Q, and even the rank of an $m \times n$ matrix are not calculated this way. There are sophisticated numerical algorithms for calculating them to a high degree of accuracy. The reader is referred to books on numerical linear algebra.

So the main virtue of Theorem 8.6.1 is that it provides a way of *constructing* an SVD for every real matrix A. In particular it shows that every real matrix A *has* a singular value decomposition[9] in the following, more general, sense:

Definition 8.9

*A **Singular Value Decomposition** (**SVD**) of an $m \times n$ matrix A of rank r is a factorization $A = U\Sigma V^T$ where U and V are orthogonal and $\Sigma = \left[\begin{array}{cc} D & 0 \\ 0 & 0 \end{array} \right]_{m \times n}$ in block form where $D = \text{diag}(d_1, d_2, \dots, d_r)$ where each $d_i > 0$, and $r \leq m$ and $r \leq n$.*

Note that for *any* SVD $A = U\Sigma V^T$ we immediately obtain some information about A:

[9] In fact every complex matrix has an SVD [J.T. Scheick, Linear Algebra with Applications, McGraw-Hill, 1997]

Lemma 8.6.3

If $A = U\Sigma V^T$ is any SVD for A as in Definition 8.9, then:

1. $r = \text{rank}\, A$.
2. *The numbers $d_1,\ d_2,\ \ldots,\ d_r$ are the singular values of A^TA in some order.*

Proof. Use the notation of Definition 8.9. We have

$$A^TA = (V\Sigma^TU^T)(U\Sigma V^T) = V(\Sigma^T\Sigma)V^T$$

so $\Sigma^T\Sigma$ and A^TA are similar $n \times n$ matrices (Definition 5.11). Hence $r = \text{rank}\, A$ by Corollary 5.4.3, proving (1.). Furthermore, $\Sigma^T\Sigma$ and A^TA have the same eigenvalues by Theorem 5.5.1; that is (using (1.)):

$$\{d_1^2,\ d_2^2,\ \ldots,\ d_r^2\} = \{\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_r\} \quad \text{are equal as sets}$$

where $\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_r$ are the positive eigenvalues of A^TA. Hence there is a permutation τ of $\{1,\ 2,\ \cdots,\ r\}$ such that $d_i^2 = \lambda_{i\tau}$ for each $i = 1,\ 2,\ \ldots,\ r$. Hence $d_i = \sqrt{\lambda_{i\tau}} = \sigma_{i\tau}$ for each i by Definition 8.7. This proves (2.). □

We note in passing that more is true. Let A be $m \times n$ of rank r, and let $A = U\Sigma V^T$ be any SVD for A. Using the proof of Lemma 8.6.3 we have $d_i = \sigma_{i\tau}$ for some permutation τ of $\{1,\ 2,\ \ldots,\ r\}$. In fact, it can be shown that there exist orthogonal matrices U_1 and V_1 obtained from U and V by τ-permuting columns and rows respectively, such that $A = U_1\Sigma_AV_1^T$ is an SVD of A.

8.6.2. Fundamental Subspaces

It turns out that any singular value decomposition contains a great deal of information about an $m \times n$ matrix A and the subspaces associated with A. For example, in addition to Lemma 8.6.3, the set $\{\mathbf{p}_1,\ \mathbf{p}_2,\ \ldots,\ \mathbf{p}_r\}$ of vectors constructed in the proof of Theorem 8.6.1 is an orthonormal basis of $\text{col}\, A$ (by (**v**) and (**viii**) in the proof). There are more such examples, which is the thrust of this subsection. In particular, there are four subspaces associated to a real $m \times n$ matrix A that have come to be called fundamental:

Definition 8.10

The ***fundamental subspaces*** *of an $m \times n$ matrix A are:*

$$\text{row}\, A = \text{span}\,\{\mathbf{x} \mid \mathbf{x} \text{ is a row of } A\}$$

$$\text{col}\, A = \text{span}\,\{\mathbf{x} \mid \mathbf{x} \text{ is a column of } A\}$$

$$\text{null}\, A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

$$\text{null}\, A^T = \{\mathbf{x} \in \mathbb{R}^n \mid A^T\mathbf{x} = \mathbf{0}\}$$

If $A = U\Sigma V^T$ is *any* SVD for the real $m \times n$ matrix A, any orthonormal bases of U and V provide orthonormal bases for each of these fundamental subspaces. We are going to prove this, but first we need three properties related to the *orthogonal complement* $U^\perp$ of a subspace U of $\mathbb{R}^n$, where (Definition 8.1):

$$U^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{x} = 0 \text{ for all } \mathbf{u} \in U\}$$

The orthogonal complement plays an important role in the Projection Theorem (Theorem 8.1.3), and we return to it in Section **??**. For now we need:

Lemma 8.6.4

If A is any matrix then:

1. $(\text{row } A)^\perp = \text{null } A$ *and* $(\text{col } A)^\perp = \text{null } A^T$.
2. *If U is any subspace of $\mathbb{R}^n$ then $U^{\perp\perp} = U$.*
3. *Let $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ be an orthonormal basis of $\mathbb{R}^m$. If $U = \text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_k\}$, then*

$$U^\perp = \text{span}\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\}$$

Proof.

1. Assume A is $m \times n$, and let $\mathbf{b}_1, \dots, \mathbf{b}_m$ be the rows of A. If $\mathbf{x}$ is a column in $\mathbb{R}^n$, then entry i of $A\mathbf{x}$ is $\mathbf{b}_i \cdot \mathbf{x}$, so $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{b}_i \cdot \mathbf{x} = 0$ for each i. Thus:

$$\mathbf{x} \in \text{null } A \quad \Leftrightarrow \quad \mathbf{b}_i \cdot \mathbf{x} = 0 \text{ for each } i \quad \Leftrightarrow \quad \mathbf{x} \in (\text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_m\})^\perp = (\text{row } A)^\perp$$

   Hence $\text{null } A = (\text{row } A)^\perp$. Now replace A by A^T to get $\text{null } A^T = (\text{row } A^T)^\perp = (\text{col } A)^\perp$, which is the other identity in (1).

2. If $\mathbf{x} \in U$ then $\mathbf{y} \cdot \mathbf{x} = 0$ for all $\mathbf{y} \in U^\perp$, that is $\mathbf{x} \in U^{\perp\perp}$. This proves that $U \subseteq U^{\perp\perp}$, so it is enough to show that $\dim U = \dim U^{\perp\perp}$. By Theorem 8.1.4 we see that $\dim V^\perp = n - \dim V$ for any subspace $V \subseteq \mathbb{R}^n$. Hence

$$\dim U^{\perp\perp} = n - \dim U^\perp = n - (n - \dim U) = \dim U, \quad \text{as required}$$

3. We have $\text{span}\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\} \subseteq U^\perp$ because $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ is orthogonal. For the other inclusion, let $\mathbf{x} \in U^\perp$ so $\mathbf{f}_i \cdot \mathbf{x} = 0$ for $i = 1, 2, \dots, k$. By the Expansion Theorem 5.3.6:

$$\begin{array}{rcccccccccccc} \mathbf{x} & = & (\mathbf{f}_1 \cdot \mathbf{x})\mathbf{f}_1 & + & \cdots & + & (\mathbf{f}_k \cdot \mathbf{x})\mathbf{f}_k & + & (\mathbf{f}_{k+1} \cdot \mathbf{x})\mathbf{f}_{k+1} & + & \cdots & + & (\mathbf{f}_m \cdot \mathbf{x})\mathbf{f}_m \\ & = & \mathbf{0} & + & \cdots & + & \mathbf{0} & + & (\mathbf{f}_{k+1} \cdot \mathbf{x})\mathbf{f}_{k+1} & + & \cdots & + & (\mathbf{f}_m \cdot \mathbf{x})\mathbf{f}_m \end{array}$$

   Hence $U^\perp \subseteq \text{span}\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\}$.

□

With this we can see how *any* SVD for a matrix A provides orthonormal bases for each of the four fundamental subspaces of A.

Theorem 8.6.2

Let A be an $m \times n$ real matrix, let $A = U\Sigma V^T$ be any SVD for A where U and V are orthogonal of size $m \times m$ and $n \times n$ respectively, and let

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}_{m \times n} \qquad \text{where} \qquad D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_r), \quad \text{with each } \lambda_i > 0$$

Write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \end{bmatrix}$ and $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r & \cdots & \mathbf{v}_n \end{bmatrix}$, so $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \dots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_r, \dots, \mathbf{v}_n\}$ are orthonormal bases of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Then

1. *$r = \operatorname{rank} A$, and the singular values of A are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_r}$.*
2. *The fundamental spaces are described as follows:*
   a. *$\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is an orthonormal basis of $\operatorname{col} A$.*
   b. *$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ is an orthonormal basis of $\operatorname{null} A^T$.*
   c. *$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is an orthonormal basis of $\operatorname{null} A$.*
   d. *$\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is an orthonormal basis of $\operatorname{row} A$.*

Proof.

1. This is Lemma 8.6.3.
2. a. As $\operatorname{col} A = \operatorname{col}(AV)$ by Lemma 5.4.3 and $AV = U\Sigma$, (a.) follows from

$$U\Sigma = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_r) & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

   b. We have $(\operatorname{col} A)^\perp \overset{(a.)}{=} (\operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\})^\perp = \operatorname{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ by Lemma 8.6.4(3). This proves (b.) because $(\operatorname{col} A)^\perp = \operatorname{null} A^T$ by Lemma 8.6.4(1).

   c. We have $\dim(\operatorname{null} A) + \dim(\operatorname{im} A) = n$ by the Dimension Theorem 7.2.4, applied to $T : \mathbb{R}^n \to \mathbb{R}^m$ where $T(\mathbf{x}) = A\mathbf{x}$. Since also $\operatorname{im} A = \operatorname{col} A$ by Lemma 8.6.1, we obtain

$$\dim(\operatorname{null} A) = n - \dim(\operatorname{col} A) = n - r = \dim(\operatorname{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\})$$

   So to prove (c.) it is enough to show that $\mathbf{v}_j \in \operatorname{null} A$ whenever $j > r$. To this end write

$$\lambda_{r+1} = \cdots = \lambda_n = 0, \qquad \text{so} \quad E^T E = \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2, \lambda_{r+1}^2, \dots, \lambda_n^2)$$

   Observe that each λ_j is an eigenvalue of $\Sigma^T\Sigma$ with eigenvector $\mathbf{e}_j =$ column j of I_n. Thus $\mathbf{v}_j = V\mathbf{e}_j$ for each j. As $A^T A = V\Sigma^T\Sigma V^T$ (proof of Lemma 8.6.3), we obtain

$$(A^T A)\mathbf{v}_j = (V\Sigma^T\Sigma V^T)(V\mathbf{e}_j) = V(\Sigma^T\Sigma\mathbf{e}_j) = V\left(\lambda_j^2\mathbf{e}_j\right) = \lambda_j^2 V\mathbf{e}_j = \lambda_j^2\mathbf{v}_j$$

   for $1 \leq j \leq n$. Thus each $\mathbf{v}_j$ is an eigenvector of $A^T A$ corresponding to λ_j^2. But then

$$\|A\mathbf{v}_j\|^2 = (A\mathbf{v}_j)^T A\mathbf{v}_j = \mathbf{v}_j^T(A^T A\mathbf{v}_j) = \mathbf{v}_j^T(\lambda_j^2\mathbf{v}_j) = \lambda_j^2\|\mathbf{v}_j\|^2 = \lambda_j^2 \quad \text{for } i = 1, \dots, n$$

   In particular, $A\mathbf{v}_j = \mathbf{0}$ whenever $j > r$, so $\mathbf{v}_j \in \operatorname{null} A$ if $j > r$, as desired. This proves (c).

d. Observe that $\text{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\} \overset{\text{(c.)}}{=} \text{null}\, A = (\text{row}\, A)^{\perp}$ by Lemma 8.6.4(1). But then parts (2) and (3) of Lemma 8.6.4 show

$$\text{row}\, A = \left((\text{row}\ A)^{\perp}\right)^{\perp} = (\text{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\})^{\perp} = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$$

This proves (d.), and hence Theorem 8.6.2.

□

Example 8.6.2

Consider the homogeneous linear system

$$A\mathbf{x} = \mathbf{0} \text{ of } m \text{ equations in } n \text{ variables}$$

Then the set of all solutions is $\text{null}\, A$. Hence if $A = U\Sigma V^T$ is any SVD for A then (in the notation of Theorem 8.6.2) $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of the set of solutions for the system. As such they are a set of **basic solutions** for the system, the most basic notion in Chapter 1.

8.6.3. The Polar Decomposition of a Real Square Matrix

If A is real and $n \times n$ the factorization in the title is related to the polar decomposition A. Unlike the SVD, in this case the decomposition is *uniquely* determined by A.

Recall (Section 8.3) that a symmetric matrix A is called positive definite if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for every column $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$. Before proceeding, we must explore the following weaker notion:

Definition 8.11

A real $n \times n$ matrix G is called ***positive***[10] *if it is symmetric and*

$$\mathbf{x}^T G\mathbf{x} \geq 0 \quad \textit{for all } \mathbf{x} \in \mathbb{R}^n$$

Clearly every positive definite matrix is positive, but the converse fails. Indeed, $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is positive because, if $\mathbf{x} = \begin{bmatrix} a & b \end{bmatrix}^T$ in $\mathbb{R}^2$, then $\mathbf{x}^T A\mathbf{x} = (a+b)^2 \geq 0$. But $\mathbf{y}^T A\mathbf{y} = 0$ if $\mathbf{y} = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$, so A is not positive definite.

Lemma 8.6.5

Let G denote an $n \times n$ positive matrix.

1. *If A is any $m \times n$ matrix and G is positive, then $A^T GA$ is positive (and $m \times m$).*

[10] Also called **positive semi-definite.**

2. *If $G = \operatorname{diag}(d_1, d_2, \cdots, d_n)$ and each $d_i \geq 0$ then G is positive.*

Proof.

1. $\mathbf{x}^T(A^T GA)\mathbf{x} = (A\mathbf{x})^T G(A\mathbf{x}) \geq 0$ because G is positive.
2. If $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$, then

$$\mathbf{x}^T G\mathbf{x} = d_1 x_1^2 + d_2 x_2^2 + \cdots + d_n x_n^2 \geq 0$$

because $d_i \geq 0$ for each i.

□

Definition 8.12

If A is a real $n \times n$ matrix, a factorization

$$A = GQ \text{ where } G \text{ is positive and } Q \text{ is orthogonal}$$

is called a ***polar decomposition*** *for A.*

Any SVD for a real square matrix A yields a polar form for A.

Theorem 8.6.3

Every square real matrix has a polar form.

Proof. Let $A = U\Sigma V^T$ be a SVD for A with Σ as in Definition 8.9 and $m = n$. Since $U^T U = I_n$ here we have

$$A = U\Sigma V^T = (U\Sigma)(U^T U)V^T = (U\Sigma U^T)(UV^T)$$

So if we write $G = U\Sigma U^T$ and $Q = UV^T$, then Q is orthogonal, and it remains to show that G is positive. But this follows from Lemma 8.6.5. □

The SVD for a square matrix A is not unique ($I_n = PI_nP^T$ for any orthogonal matrix P). But given the proof of Theorem 8.6.3 it is surprising that the polar decomposition *is* unique.[11] We omit the proof.

The name "polar form" is reminiscent of the same form for complex numbers (see Appendix **??**). This is no coincidence. To see why, we represent the complex numbers as real 2×2 matrices. Write $\mathbf{M}_2(\mathbb{R})$ for the set of all real 2×2 matrices, and define

$$\sigma : \mathbb{C} \to \mathbf{M}_2(\mathbb{R}) \quad \text{by} \quad \sigma(a + bi) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \text{ for all } a + bi \text{ in } \mathbb{C}$$

[11] See J.T. Scheick, Linear Algebra with Applications, McGraw-Hill, 1997, page 379.

One verifies that σ preserves addition and multiplication in the sense that

$$\sigma(zw) = \sigma(z)\sigma(w) \qquad \text{and} \qquad \sigma(z+w) = \sigma(z) + \sigma(w)$$

for all complex numbers z and w. Since θ is one-to-one we may *identify* each complex number $a+bi$ with the matrix $\theta(a+bi)$, that is we write

$$a+bi = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \quad \text{for all } a+bi \text{ in } \mathbb{C}$$

Thus $0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$, $i = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, and $r = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ if r is real.

If $z = a+bi$ is nonzero then the *absolute value* $r = |z| = \sqrt{a^2+b^2} \neq 0$. If θ is the *angle* of z in standard position, then $\cos\theta = a/r$ and $\sin\theta = b/r$. Observe:

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = GQ \qquad \textbf{(xiii)}$$

where $G = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ is positive and $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal. But in $\mathbb{C}$ we have $G = r$ and $Q = \cos\theta + i\sin\theta$ so (**xiii**) reads $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$ which is the *classical polar form* for the complex number $a+bi$. This is why (**xiii**) is called the polar form of the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$; Definition 8.12 simply adopts the terminology for $n \times n$ matrices.

8.6.4. The Pseudoinverse of a Matrix

It is impossible for a non-square matrix A to have an inverse (see the footnote to Definition 2.11). Nonetheless, one candidate for an "inverse" of A is an $m \times n$ matrix B such that

$$ABA = A \qquad \text{and} \qquad BAB = B$$

Such a matrix B is called a *middle inverse* for A. If A is invertible then A^{-1} is the unique middle inverse for A, but a middle inverse is not unique in general, even for square matrices. For example, if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ then $B = \begin{bmatrix} 1 & 0 & 0 \\ b & 0 & 0 \end{bmatrix}$ is a middle inverse for A for any b.

If $ABA = A$ and $BAB = B$ it is easy to see that AB and BA are both idempotent matrices. In 1955 Roger Penrose observed that the middle inverse is unique if both AB and BA are symmetric. We omit the proof.

Theorem 8.6.4: Penrose' Theorem[12]

Given any real $m \times n$ matrix A, there is exactly one $n \times m$ matrix B such that A and B satisfy the following conditions:

P1 *$ABA = A$ and $BAB = B$.*

P2 *Both AB and BA are symmetric.*

Definition 8.13

*Let A be a real $m \times n$ matrix. The **pseudoinverse** of A is the unique $n \times m$ matrix A^+ such that A and A^+ satisfy **P1** and **P2**, that is:*

$$AA^+A = A, \qquad A^+AA^+ = A^+, \qquad \text{and both } AA^+ \text{ and } A^+A \text{ are symmetric}$$ [13]

If A is invertible then $A^+ = A^{-1}$ as expected. In general, the symmetry in conditions P1 and P2 shows that A is the pseudoinverse of A^+, that is $A^{++} = A$.

[12] R. Penrose, *A generalized inverse for matrices*, Proceedings of the Cambridge Philosophical Society **51** (1955), 406-413. In fact Penrose proved this for any complex matrix, where AB and BA are both required to be hermitian (see Definition **??** in the following section).

[13] Penrose called the matrix A^+ the generalized inverse of A, but the term pseudoinverse is now commonly used. The matrix A^+ is also called the **Moore-Penrose** inverse after E.H. Moore who had the idea in 1935 as part of a larger work on "General Analysis". Penrose independently re-discovered it 20 years later.

Theorem 8.6.5

Let A be an $m \times n$ matrix.

1. *If* $\text{rank}\, A = m$ *then* AA^T *is invertible and* $A^+ = A^T(AA^T)^{-1}$.
2. *If* $\text{rank}\, A = n$ *then* A^TA *is invertible and* $A^+ = (A^TA)^{-1}A^T$.

Proof. Here AA^T (respectively A^TA) is invertible by Theorem 5.4.4 (respectively Theorem 5.4.3). The rest is a routine verification. □

In general, given an $m \times n$ matrix A, the pseudoinverse A^+ can be computed from any SVD for A. To see how, we need some notation. Let $A = U\Sigma V^T$ be an SVD for A (as in Definition 8.9) where U and V are orthogonal and $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$ in block form where $D = \text{diag}\,(d_1,\, d_2,\, \dots,\, d_r)$ where each $d_i > 0$. Hence D is invertible, so we make:

Definition 8.14

$$\Sigma' = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{n \times m}.$$

A routine calculation gives:

Lemma 8.6.6

- $\Sigma\Sigma'\Sigma = \Sigma$
- $\Sigma\Sigma' = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times m}$
- $\Sigma'\Sigma\Sigma' = \Sigma'$
- $\Sigma'\Sigma = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times n}$

That is, Σ' is the pseudoinverse of Σ.

Now given $A = U\Sigma V^T$, define $B = V\Sigma' U^T$. Then

$$ABA = (U\Sigma V^T)(V\Sigma' U^T)(U\Sigma V^T) = U(\Sigma\Sigma'\Sigma)V^T = U\Sigma V^T = A$$

by Lemma 8.6.6. Similarly $BAB = B$. Moreover $AB = U(\Sigma\Sigma')U^T$ and $BA = V(\Sigma'\Sigma)V^T$ are both symmetric again by Lemma 8.6.6. This proves

Theorem 8.6.6

Let A be real and $m \times n$, and let $A = U\Sigma V^T$ is any SVD for A as in Definition 8.9. Then $A^+ = V\Sigma' U^T$.

Of course we can always use the SVD constructed in Theorem 8.6.1 to find the pseudoinverse. If $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, we observed above that $B = \begin{bmatrix} 1 & 0 & 0 \\ b & 0 & 0 \end{bmatrix}$ is a middle inverse for A for any b. Furthermore AB is symmetric but BA is not, so $B \neq A^+$.

Example 8.6.3

Find A^+ if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution. $A^TA = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0$ and corresponding eigenvectors $\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{q}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Hence $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = I_2$. Also A has rank 1 with singular values $\sigma_1 = 1$ and $\sigma_2 = 0$, so $\Sigma_A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = A$ and $\Sigma_A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = A^T$ in this case.

Since $A\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $A\mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, we have $\mathbf{p}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ which extends to an orthonormal basis $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ of $\mathbb{R}^3$ where (say) $\mathbf{p}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{p}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Hence $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} = I$, so the SVD for A is $A = P\Sigma_A Q^T$. Finally, the pseudoinverse of A is $A^+ = Q\Sigma_A' P^T = \Sigma_A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Note that $A^+ = A^T$ in this case.

The following Lemma collects some properties of the pseudoinverse that mimic those of the inverse. The verifications are left as exercises.

Lemma 8.6.7

Let A be an $m \times n$ matrix of rank r.

1. $A^{++} = A$.
2. *If A is invertible then* $A^+ = A^{-1}$.
3. $(A^T)^+ = (A^+)^T$.
4. $(kA)^+ = kA^+$ *for any real* k.
5. $(UAV)^+ = U^T(A^+)V^T$ *whenever U and V are orthogonal.*

Exercises for 8.6

Exercise 8.6.1 If $ACA = A$ show that $B = CAC$ is a middle inverse for A.

Exercise 8.6.2 For any matrix A show that

$$\Sigma_{A^T} = (\Sigma_A)^T$$

Exercise 8.6.3 If A is $m \times n$ with all singular values positive, what is $\text{rank}\, A$?

Exercise 8.6.4 If A has singular values $\sigma_1, \dots, \sigma_r$, what are the singular values of:

a) A^T

b) tA where $t > 0$ is real

c) A^{-1} assuming A is invertible.

b. $t\sigma_1, \dots, t\sigma_r$.

Exercise 8.6.5 If A is square show that $\det A$ is the product of the singular values of A.

Exercise 8.6.6 If A is square and real, show that $A = 0$ if and only if every eigenvalue of A is 0.

Exercise 8.6.7 Given a SVD for an invertible matrix A, find one for A^{-1}. How are Σ_A and $\Sigma_{A^{-1}}$ related?
If $A = U\Sigma V^T$ then Σ is invertible, so $A^{-1} = V\Sigma^{-1}U^T$ is a SVD.

Exercise 8.6.8 Let $A^{-1} = A = A^T$ where A is $n \times n$. Given any orthogonal $n \times n$ matrix U, find an orthogonal matrix V such that $A = U\Sigma_A V^T$ is an SVD for A. If $A = \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \end{array}\right]$ do this for:

a) $U = \frac{1}{5}\left[\begin{array}{rr} 3 & -4 \\ 4 & 3 \end{array}\right]$

b) $U = \frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & -1 \\ 1 & 1 \end{array}\right]$

b. First $A^T A = I_n$ so $\Sigma_A = I_n$.

$$\begin{aligned} A &= \frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & 1 \\ 1 & -1 \end{array}\right]\left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right]\frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & 1 \\ -1 & 1 \end{array}\right] \\ &= \frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & -1 \\ 1 & 1 \end{array}\right]\frac{1}{\sqrt{2}}\left[\begin{array}{rr} -1 & 1 \\ 1 & 1 \end{array}\right] \\ &= \left[\begin{array}{rr} -1 & 0 \\ 0 & 1 \end{array}\right] \end{aligned}$$

Exercise 8.6.9 Find a SVD for the following matrices:

a) $A = \left[\begin{array}{rr} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{array}\right]$

b) $\left[\begin{array}{rrr} 1 & 1 & 1 \\ -1 & 0 & -2 \\ 1 & 2 & 0 \end{array}\right]$

b.

$$A = F$$

$$= \frac{1}{5}\left[\begin{array}{rr} 3 & 4 \\ 4 & -3 \end{array}\right]\left[\begin{array}{rrrr} 20 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \end{array}\right]\frac{1}{2}\left[\begin{array}{rrrr} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{array}\right]$$

Exercise 8.6.10 Find an SVD for $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array}\right]$.

Exercise 8.6.11 If $A = U\Sigma V^T$ is an SVD for A, find an SVD for A^T.

Exercise 8.6.12 Let A be a real, $m \times n$ matrix with positive singular values $\sigma_1, \sigma_2, \dots, \sigma_r$, and write

$$s(x) = (x - \sigma_1)(x - \sigma_2)\cdots(x - \sigma_r)$$

a. Show that $c_{A^TA}(x) = s(x)x^{n-r}$ and $c_{A^TA}(c) = s(x)x^{m-r}$.

b. If $m \leq n$ conclude that $c_{A^TA}(x) = s(x)x^{n-m}$.

Exercise 8.6.13 If G is positive show that:

a. rG is positive if $r \geq 0$

b. $G + H$ is positive for any positive H.

b. If $\mathbf{x} \in \mathbb{R}^n$ then $\mathbf{x}^T(G+H)\mathbf{x} = \mathbf{x}^T G\mathbf{x} + \mathbf{x}^T H\mathbf{x} \geq 0 + 0 = 0$.

Exercise 8.6.14 If G is positive and λ is an eigenvalue, show that $\lambda \geq 0$.

Exercise 8.6.15 If G is positive show that $G = H^2$ for some positive matrix H. [Hint: Preceding exercise and Lemma 8.6.5]

Exercise 8.6.16 If A is $n \times n$ show that AA^T and A^TA are similar. [Hint: Start with an SVD for A.]

Exercise 8.6.17 Find A^+ if:

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

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

b. $\begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} \\ -\frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}$

Exercise 8.6.18 Show that $(A^+)^T = (A^T)^+$.