6.1 Introduction to eigen value (*Refer 4.2 for applications)

Another consequence of Theorem 4.14 is that any vector space $\mathcal{V}$ having a finite spanning set S must be finite dimensional. This is because a maximal linearly independent subset of S, which must also be finite, is a basis for $\mathcal{V}$ (see Exercise 24).

We also have the following result for spanning sets:

Theorem 4.15 Let $\mathcal{V}$ be a vector space, and let B be a minimal spanning set for $\mathcal{V}$. Then B is a basis for $\mathcal{V}$.

The phrase "B is a **minimal spanning set** for $\mathcal{V}$" means that both of the following are true:

- B is a subset of $\mathcal{V}$ that spans $\mathcal{V}$.
- If $C \subset B$ and $C \neq B$, then C does not span $\mathcal{V}$.

The converse of Theorem 4.15 is true as well (see Exercise 21).

Example 14

Consider the subsets S and B of $\mathbb{R}^3$ given in Example 13. We can use Theorem 4.15 to give another justification that B is a basis for $\mathcal{V} = \text{span}(S)$. Recall from Example 13 that every vector in S is a linear combination of vectors in B, so $S \subseteq \text{span}(B)$. This fact along with $B \subseteq S$ and Corollary 4.6 shows that $\text{span}(B) = \text{span}(S) = \mathcal{V}$. Also, neither vector in B is a scalar multiple of the other, so that neither vector alone can span $\mathcal{V}$ (why?). Hence, B is a minimal spanning set for $\mathcal{V}$, and by Theorem 4.15, B is a basis for $\text{span}(S)$. ■

Dimension of a Subspace

We conclude this section with the result that every subspace of a finite dimensional vector space is also finite dimensional. This is important because it tells us that the theorems we have developed about finite dimension apply to all *subspaces* of our basic examples $\mathbb{R}^n$, $\mathcal{M}_{mn}$, and $\mathcal{P}_n$.

Theorem 4.16 Let $\mathcal{V}$ be a finite dimensional vector space, and let $\mathcal{W}$ be a subspace of $\mathcal{V}$. Then $\mathcal{W}$ is also finite dimensional with $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$. Moreover, $\dim(\mathcal{W}) = \dim(\mathcal{V})$ if and only if $\mathcal{W} = \mathcal{V}$.

The proof of Theorem 4.16 is left for you to do, with hints, in Exercise 22. The only subtle part of this proof involves showing that $\mathcal{W}$ actually has a basis.[4]

[4] Although it is true that *every* vector space has a basis, we must be careful here, because we have not proven this. In fact, Theorem 4.16 establishes that every subspace of a finite dimensional vector space *does* have a basis and that this basis is finite. Although every finite dimensional vector space has a finite basis by definition, the proof that every infinite dimensional vector space has a basis requires advanced set theory and is beyond the scope of this text.

Example 15

Consider the nested sequence of subspaces of $\mathbb{R}^3$ given by $\{\mathbf{0}\} \subset \{\text{scalar multiples of } [4,-7,0]\} \subset$ xy-plane $\subset \mathbb{R}^3$. Their respective dimensions are $0, 1, 2$, and 3 (why?). Hence, the dimensions of each successive pair of these subspaces satisfy the inequality given in Theorem 4.16. ■

Example 16

It can be shown that $B = \{x^3 + 2x^2 - 4x + 18, 3x^2 + 4x - 4, x^3 + 5x^2 - 3, 3x + 2\}$ is a linearly independent subset of $\mathcal{P}_3$. Therefore, by part (2) of Theorem 4.13, B is a basis for $\mathcal{P}_3$. However, we can also reach the same conclusion from Theorem 4.16. For, $\mathcal{W} = \text{span}(B)$ has B as a basis (why?), and hence, $\dim(\mathcal{W}) = 4$. But since $\mathcal{W}$ is a subspace of $\mathcal{P}_3$ and $\dim(\mathcal{P}_3) = 4$, Theorem 4.16 implies that $\mathcal{W} = \mathcal{P}_3$. Hence, B is a basis for $\mathcal{P}_3$. ■

New Vocabulary

basis
dimension
finite dimensional (vector space)
infinite dimensional (vector space)
maximal linearly independent set
minimal spanning set
standard basis (for $\mathbb{R}^n$, $\mathcal{M}_{mn}$, $\mathcal{P}_n$)

Highlights

- A basis is a subset of a vector space that both spans and is linearly independent.
- If a finite basis exists for a vector space, the vector space is said to be finite dimensional.
- For a finite dimensional vector space, all bases have the same number of vectors, and this number is known as the dimension of the vector space.
- The standard basis for $\mathbb{R}^n$ is $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$; $\dim(\mathbb{R}^n) = n$.
- The standard basis for $\mathcal{P}_n$ is $\{1, x, x^2, \ldots, x^n\}$; $\dim(\mathcal{P}_n) = n + 1$.
- The standard basis for $\mathcal{M}_{mn}$ is $\{\Psi_{ij}\}$, where each Ψ_{ij} has a 1 in the (i,j) entry, and zeroes elsewhere; $\dim(\mathcal{M}_{mn}) = m \cdot n$.
- The basis for the trivial vector space $\{\mathbf{0}\}$ is the empty set $\{\ \}$; $\dim(\{\mathbf{0}\}) = 0$.
- If no finite basis exists for a vector space, the vector space is said to be infinite dimensional. $\mathcal{P}$ is an infinite dimensional vector space, as is the set of all real-valued functions (under normal operations).
- In a vector space $\mathcal{V}$ with dimension n, the size of a spanning set S is always $\geq n$. If $|S| = n$, then S is a basis for $\mathcal{V}$.
- In a vector space $\mathcal{V}$ with dimension n, the size of a linearly independent set T is always $\leq n$. If $|T| = n$, then T is a basis for $\mathcal{V}$.

- A maximal linearly independent set in a vector space is a basis.
- A minimal spanning set in a vector space is a basis.
- In a vector space $\mathcal{V}$ with dimension n, the dimension of a subspace $\mathcal{W}$ is always $\leq n$. If $\dim(\mathcal{W}) = n$, then $\mathcal{W} = \mathcal{V}$.

EXERCISES FOR SECTION 4.5

1. Prove that each of the following subsets of $\mathbb{R}^4$ is a basis for $\mathbb{R}^4$ by showing both that it spans $\mathbb{R}^4$ and that is linearly independent:

(a) $\{[2,1,0,0],[0,1,1,-1],[0,-1,2,-2],[3,1,0,-2]\}$

(b) $\{[6,1,1,-1],[1,0,0,9],[-2,3,2,4],[2,2,5,-5]\}$

(c) $\{[1,1,1,1],[1,1,1,-1],[1,1,-1,-1],[1,-1,-1,-1]\}$

(d) $\left\{\left[\frac{15}{2},5,\frac{12}{5},1\right],\left[2,\frac{1}{2},\frac{3}{4},1\right],\left[-\frac{13}{2},1,0,4\right],\left[\frac{18}{5},0,\frac{1}{5},-\frac{1}{5}\right]\right\}$

2. Prove that the following set is a basis for $\mathcal{M}_{22}$ by showing that it spans $\mathcal{M}_{22}$ and is linearly independent:

$$\left\{\begin{bmatrix}1 & 4\\ 2 & 0\end{bmatrix},\begin{bmatrix}0 & 2\\ 1 & 0\end{bmatrix},\begin{bmatrix}-3 & 1\\ -1 & 0\end{bmatrix},\begin{bmatrix}5 & -2\\ 0 & -3\end{bmatrix}\right\}.$$

3. Show that the subset $\left\{x^4,\ x^4-x^3,\ x^4-x^3+x^2,\ x^4-x^3+x^2-x,\ x^3-1\right\}$ of $\mathcal{P}_4$ is a basis for $\mathcal{P}_4$.

4. Determine which of the following subsets of $\mathbb{R}^4$ form a basis for $\mathbb{R}^4$:

★**(a)** $S = \{[7,1,2,0],[8,0,1,-1],[1,0,0,-2]\}$

(b) $S = \{[1,3,2,0],[-2,0,6,7],[0,6,10,7]\}$

★**(c)** $S = \{[7,1,2,0],[8,0,1,-1],[1,0,0,-2],[3,0,1,-1]\}$

(d) $S = \{[1,3,2,0],[-2,0,6,7],[0,6,10,7],[2,10,-3,1]\}$

★**(e)** $S = \{[1,2,3,2],[1,4,9,3],[6,-2,1,4],[3,1,2,1],[10,-9,-15,6]\}$

5. **(a)** Show that $B = \{[2,3,0,-1],[-1,1,1,-1]\}$ is a maximal linearly independent subset of $S = \{[1,4,1,-2],[-1,1,1,-1],[3,2,-1,0],[2,3,0,-1]\}$.

★**(b)** Calculate $\dim(\text{span}(S))$.

★**(c)** Does $\text{span}(S) = \mathbb{R}^4$? Why or why not?

(d) Is B a minimal spanning set for $\text{span}(S)$? Why or why not?

6. **(a)** Show that $B = \{x^3-x^2+2x+1,\ 2x^3+4x-7,\ 3x^3-x^2-6x+6\}$ is a maximal linearly independent subset of $S = \{x^3-x^2+2x+1,\ x-1,\ 2x^3+4x-7, x^3-3x^2-22x+34, 3x^3-x^2-6x+6\}$.

(b) Calculate $\dim(\text{span}(S))$.

(c) Does $\operatorname{span}(S) = \mathcal{P}_3$? Why or why not?

(d) Is B a minimal spanning set for $\operatorname{span}(S)$? Why or why not?

7. Let $\mathcal{W}$ be the solution set to the matrix equation $\mathbf{AX} = \mathbf{O}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 0 & -1 \\ 2 & -1 & 0 & 1 & 3 \\ 1 & -3 & -1 & 1 & 4 \\ 2 & 9 & 4 & -1 & -7 \end{bmatrix}.$$

(a) Show that $\mathcal{W}$ is a subspace of $\mathbb{R}^5$.

(b) Find a basis for $\mathcal{W}$.

(c) Show that $\dim(\mathcal{W}) + \operatorname{rank}(\mathbf{A}) = 5$.

8. Prove that every proper nontrivial subspace of $\mathbb{R}^3$ can be thought of, from a geometric point of view, as either a line through the origin or a plane through the origin.

9. Let $\mathbf{f}$ be a polynomial of degree n. Show that the set $\{\mathbf{f}, \mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}\}$ is a basis for $\mathcal{P}_n$ (where $\mathbf{f}^{(i)}$ denotes the ith derivative of $\mathbf{f}$). (Hint: See Exercise 23 in Section 4.4.)

10. **(a)** Let $\mathbf{A}$ be a 2×2 matrix. Prove that there are real numbers $a_0, a_1, \ldots, a_4$, not all zero, such that $a_4\mathbf{A}^4 + a_3\mathbf{A}^3 + a_2\mathbf{A}^2 + a_1\mathbf{A} + a_0\mathbf{I}_2 = \mathbf{O}_2$. (Hint: You can assume that $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4$, and $\mathbf{I}_2$ are all distinct because if they are not, opposite nonzero coefficients can be chosen for any identical pair to demonstrate that the given statement holds.)

(b) Suppose $\mathbf{B}$ is an $n \times n$ matrix. Show that there must be a nonzero polynomial $\mathbf{p} \in \mathcal{P}_{n^2}$ such that $\mathbf{p}(\mathbf{B}) = \mathbf{O}_n$.

11. **(a)** Show that $B = \{(x-2), x(x-2), x^2(x-2), x^3(x-2), x^4(x-2)\}$ is a basis for $\mathcal{V} = \{\mathbf{p} \in \mathcal{P}_5 | \mathbf{p}(2) = 0\}$.

★(b) What is $\dim(\mathcal{V})$?

★(c) Find a basis for $\mathcal{W} = \{\mathbf{p} \in \mathcal{P}_5 \mid \mathbf{p}(2) = \mathbf{p}(3) = 0\}$.

★(d) Calculate $\dim(\mathcal{W})$.

★12. Let $\mathcal{V}$ be a finite dimensional vector space.

(a) Let S be a subset of $\mathcal{V}$ with $\dim(\mathcal{V}) \leq |S|$. Find an example to show that S need not span $\mathcal{V}$.

(b) Let T be a subset of $\mathcal{V}$ with $|T| \leq \dim(\mathcal{V})$. Find an example to show that T need not be linearly independent.

13. Let S be a subset of a finite dimensional vector space $\mathcal{V}$ such that $|S| = \dim(\mathcal{V})$. If S is not a basis for $\mathcal{V}$, prove that S neither spans $\mathcal{V}$ nor is linearly independent.

14. Let $\mathcal{V}$ be an n-dimensional vector space, and let S be a subset of $\mathcal{V}$ containing exactly n elements. Prove that S spans $\mathcal{V}$ if and only if S is linearly independent.

15. Let $\mathbf{A}$ be a nonsingular $n \times n$ matrix, and let B be a basis for $\mathbb{R}^n$.

(a) Show that $B_1 = \{\mathbf{Av} \mid \mathbf{v} \in B\}$ is also a basis for $\mathbb{R}^n$. (Treat the vectors in B as column vectors.)

(b) Show that $B_2 = \{\mathbf{vA} \mid \mathbf{v} \in B\}$ is also a basis for $\mathbb{R}^n$. (Treat the vectors in B as row vectors.)

(c) Letting B be the standard basis for $\mathbb{R}^n$, use the result of part (a) to show that the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

(d) Prove that the rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

16. Prove that $\mathcal{P}$ is infinite dimensional by showing that no finite subset S of $\mathcal{P}$ can span $\mathcal{P}$, as follows:

(a) Let S be a finite subset of $\mathcal{P}$. Show that $S \subseteq \mathcal{P}_n$, for some n.

(b) Use part (a) to prove that $\text{span}(S) \subseteq \mathcal{P}_n$.

(c) Conclude that S cannot span $\mathcal{P}$.

17. **(a)** Prove that if a vector space $\mathcal{V}$ has an infinite linearly independent subset, then $\mathcal{V}$ is not finite dimensional.

(b) Use part (a) to prove that any vector space having $\mathcal{P}$ as a subspace is not finite dimensional.

18. The purpose of this exercise is to prove Theorem 4.14. Let $\mathcal{V}$, S, and B be as given in the statement of the theorem. Suppose $B \neq S$, and $\mathbf{w} \in S$ with $\mathbf{w} \notin B$.

(a) Explain why it is sufficient to prove that B spans $\mathcal{V}$.

▶**(b)** Prove that if $S \subseteq \text{span}(B)$, then B spans $\mathcal{V}$.

▶**(c)** Let $C = B \cup \{\mathbf{w}\}$. Prove that C is linearly dependent.

(d) Use part (c) to prove that $\mathbf{w} \in \text{span}(B)$. (Also see part (a) of Exercise 26 in Section 4.4.)

(e) Tie together all parts to finish the proof.

19. The purpose of this exercise is to prove Theorem 4.15.

(a) Explain why it is sufficient to prove the following statement: Let S be a spanning set for a vector space $\mathcal{V}$. If S is a minimal spanning set for $\mathcal{V}$, then S is linearly independent.

▶**(b)** State the contrapositive of the statement in part (a).

▶**(c)** Prove the statement from part (b). (Hint: Use Exercise 12 from Section 4.4.)

20. Let B be a basis for a vector space $\mathcal{V}$. Prove that B is a maximal linearly independent subset of $\mathcal{V}$. (Note: You may *not* use $\dim(\mathcal{V})$ in your proof, since $\mathcal{V}$ could be infinite dimensional.)

21. Let B be a basis for a vector space $\mathcal{V}$. Prove that B is a minimal spanning set for $\mathcal{V}$. (Note: You may *not* use $\dim(\mathcal{V})$ in your proof, since $\mathcal{V}$ could be infinite dimensional.)

22. The purpose of this exercise is to prove Theorem 4.16. Let $\mathcal{V}$ and $\mathcal{W}$ be as given in the theorem. Consider the set A of nonnegative integers defined by $A = \{k \mid \text{a set } T \text{ exists with } T \subseteq \mathcal{W}, |T| = k, \text{and } T \text{ linearly independent}\}$.

(a) Prove that $0 \in A$. (Hence, A is nonempty.)

(b) Prove that $k \in A$ implies $k \leq \dim(\mathcal{V})$. (Hint: Use Theorem 4.13.) (Hence, A is finite.)

▶**(c)** Let n be the largest element of A. Let T be a linearly independent subset of $\mathcal{W}$ such that $|T| = n$. Prove T is a maximal linearly independent subset of $\mathcal{W}$.

▶**(d)** Use part (c) and Theorem 4.14 to prove that T is a basis for $\mathcal{W}$.

(e) Conclude that $\mathcal{W}$ is finite dimensional and use part (b) to show $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$.

(f) Prove that if $\dim(\mathcal{W}) = \dim(\mathcal{V})$, then $\mathcal{W} = \mathcal{V}$. (Hint: Let T be a basis for $\mathcal{W}$ and use part (2) of Theorem 4.13 to show that T is also a basis for $\mathcal{V}$.)

(g) Prove the converse of part (f).

23. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^n$ with $\dim(\mathcal{V}) = n - 1$. (Such a subspace is called a **hyperplane** in $\mathbb{R}^n$.) Prove that there is a nonzero $\mathbf{x} \in \mathbb{R}^n$ such that $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{v} = 0\}$. (Hint: Set up a homogeneous system of equations whose coefficient matrix has a basis for $\mathcal{V}$ as its rows. Then notice that this $(n-1) \times n$ system has at least one nontrivial solution, say $\mathbf{x}$.)

24. Let $\mathcal{V}$ be a vector space and let S be a finite spanning set for $\mathcal{V}$. Prove that $\mathcal{V}$ is finite dimensional.

★**25.** True or False:

(a) A set B of vectors in a vector space $\mathcal{V}$ is a basis for $\mathcal{V}$ if B spans $\mathcal{V}$ and B is linearly independent.

(b) All bases for $\mathcal{P}_4$ have four elements.

(c) $\dim(\mathcal{M}_{43}) = 7$.

(d) If S is a spanning set for $\mathcal{W}$ and $\dim(\mathcal{W}) = n$, then $|S| \leq n$.

(e) If T is a linearly independent set in $\mathcal{W}$ and $\dim(\mathcal{W}) = n$, then $|T| = n$.

(f) If T is a linearly independent set in a finite dimensional vector space $\mathcal{W}$ and S is a finite spanning set for $\mathcal{W}$, then $|T| \leq |S|$.

(g) If $\mathcal{W}$ is a subspace of a finite dimensional vector space $\mathcal{V}$, then $\dim(\mathcal{W}) < \dim(\mathcal{V})$.

(h) Every subspace of an infinite dimensional vector space is infinite dimensional.

(i) If T is a maximal linearly independent set for a vector space $\mathcal{V}$ and S is a minimal spanning set for $\mathcal{V}$, then $S = T$.

(j) If $\mathbf{A}$ is a nonsingular 4×4 matrix, then the rows of $\mathbf{A}$ are a basis for $\mathbb{R}^4$.

4.6 CONSTRUCTING SPECIAL BASES

In this section, we present additional methods for finding a basis for a given finite dimensional vector space, starting with either a spanning set or a linearly independent subset.

Using Row Reduction to Construct a Basis

Recall the Simplified Span Method from Section 4.3. Using that method, we were able to simplify the form of span(S) for a subset S of $\mathbb{R}^n$. This was done by creating a matrix $\mathbf{A}$ whose rows are the vectors in S, and then row reducing $\mathbf{A}$ to obtain a reduced row echelon form matrix $\mathbf{C}$. We discovered that a simplified form of span(S) is given by the set of all linear combinations of the nonzero rows of $\mathbf{C}$. Now, each nonzero row of the matrix $\mathbf{C}$ has a (pivot) 1 in a column in which all other rows have zeroes, so the nonzero rows of $\mathbf{C}$ must be linearly independent. Thus, the nonzero rows of $\mathbf{C}$ not only span S but are linearly independent as well, and so they form a basis for span(S). Therefore, whenever we use the Simplified Span Method on a subset S of $\mathbb{R}^n$, we are actually creating a basis for span(S).

Example 1

Let $S = \{[2,-2,3,5,5], [-1,1,4,14,-8], [4,-4,-2,-14,18], [3,-3,-1,-9,13]\}$, a subset of $\mathbb{R}^5$. We can use the Simplified Span Method to find a basis B for $\mathcal{V} = \text{span}(S)$. We construct the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & 3 & 5 & 5 \\ -1 & 1 & 4 & 14 & -8 \\ 4 & -4 & -2 & -14 & 18 \\ 3 & -3 & -1 & -9 & 13 \end{bmatrix},$$

whose rows are the vectors in S. The reduced row echelon form matrix for $\mathbf{A}$ is

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & -2 & 4 \\ 0 & 0 & 1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore, the desired basis for $\mathcal{V}$ is the set $B = \{[1,-1,0,-2,4], [0,0,1,3,-1]\}$ of nonzero rows of $\mathbf{C}$, and $\dim(\mathcal{V}) = 2$. ■

In general, the Simplified Span Method creates a basis of vectors with a simpler form than the original vectors. This is because a reduced row echelon form matrix has the simplest form of all matrices that are row equivalent to it.

This method can also be adapted to vector spaces other than $\mathbb{R}^n$, as in the next example.

Example 2

Consider the subset $\mathcal{S} = \{x^2 - 3x + 5, 3x^3 + 4x - 8, 6x^3 - x^2 + 11x - 21, 2x^5 - 7x^3 + 5x\}$ of $\mathcal{P}_5$. We use the Simplified Span Method to find a basis for $\mathcal{W} = \text{span}(\mathcal{S})$.

Since $\mathcal{S}$ is a subset of $\mathcal{P}_5$ instead of $\mathbb{R}^n$, we must alter our method slightly. We cannot use the polynomials in $\mathcal{S}$ themselves as rows of a matrix, so we "peel off" their coefficients to create four 6-vectors, which we use as the rows of the following matrix:

$$\mathbf{A} = \begin{array}{c} \begin{array}{cccccc} x^5 & x^4 & x^3 & x^2 & x & 1 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 0 & 1 & -3 & 5 \\ 0 & 0 & 3 & 0 & 4 & -8 \\ 0 & 0 & 6 & -1 & 11 & -21 \\ 2 & 0 & -7 & 0 & 5 & 0 \end{array}\right] \end{array}.$$

Row reducing this matrix produces

$$\mathbf{C} = \begin{array}{c} \begin{array}{cccccc} x^5 & x^4 & x^3 & x^2 & x & 1 \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & \frac{43}{6} & -\frac{28}{3} \\ 0 & 0 & 1 & 0 & \frac{4}{3} & -\frac{8}{3} \\ 0 & 0 & 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}.$$

The nonzero rows of $\mathbf{C}$ yield the following three-element basis for $\mathcal{W}$:

$$D = \left\{x^5 + \frac{43}{6}x - \frac{28}{3},\ x^3 + \frac{4}{3}x - \frac{8}{3},\ x^2 - 3x + 5\right\}.$$

Hence, $\dim(\mathcal{W}) = 3$. ■

Every Spanning Set for a Finite Dimensional Vector Space Contains a Basis

Sometimes, we are interested in reducing a spanning set to a basis by eliminating redundant vectors without changing the form of the original vectors. The next theorem asserts that this is possible; that is, if $\mathcal{V}$ is a finite dimensional vector space, then any spanning set of $\mathcal{V}$, finite or infinite, must contain a basis for $\mathcal{V}$.

Theorem 4.17 If $\mathcal{S}$ is a spanning set for a finite dimensional vector space $\mathcal{V}$, then there is a set $B \subseteq \mathcal{S}$ that is a basis for $\mathcal{V}$.

The proof of this theorem is very similar to the first part of the proof of Theorem 4.16[5] and is left as Exercise 14.

Example 3

Let $S = \{[1,3,-2],[2,1,4],[0,5,-8],[1,-7,14]\}$, and let $\mathcal{V} = \text{span}(S)$. Theorem 4.17 indicates that some subset of S is a basis for $\mathcal{V}$. Now, the equations

$$[0,5,-8] = 2[1,3,-2] - [2,1,4] \quad \text{and}$$
$$[1,-7,14] = -3[1,3,-2] + 2[2,1,4]$$

show that the subset $B = \{[1,3,-2],[2,1,4]\}$ is a maximal linearly independent subset of S (why?). Hence, by Theorem 4.14, B is a basis for $\mathcal{V}$ contained in S. ■

Shrinking a Spanning Set to a Basis Using Row Reduction

As Example 3 illustrates, to find a subset B of a spanning set S that is a basis for span(S), it is necessary to remove enough redundant vectors from S until we are left with a (maximal) linearly independent subset of S. This can be done using the Independence Test Method from Section 4.4. Suppose we row reduce the matrix whose columns are all the vectors in S. Then those vectors of S corresponding to the pivot columns form a linearly independent subset B. This is because if we had row reduced the matrix having just these columns, every column would have had a pivot. Also, no larger subset of S containing B can be linearly independent because reinserting a column corresponding to any of the remaining vectors would result in a nonpivot column after row reduction. Therefore, B is a maximal linearly independent subset of S, and hence is a basis for span(S). This procedure is illustrated in the next two examples.

Example 4

Consider the subset $S = \{[1,2,-1],[3,6,-3],[4,1,2],[0,0,0],[-1,5,-5]\}$ of $\mathbb{R}^3$. We use the Independence Test Method to find a subset B of S that is a basis for $\mathcal{V} = \text{span}(S)$. We form the matrix $\mathbf{A}$ whose columns are the vectors in S, and then row reduce

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 & 0 & -1 \\ 2 & 6 & 1 & 0 & 5 \\ -1 & -3 & 2 & 0 & -5 \end{bmatrix} \quad \text{to obtain} \quad \mathbf{C} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since there are nonzero pivots in the first and third columns of $\mathbf{C}$, we choose $B = \{[1,2,-1],[4,1,2]\}$, the first and third vectors in S. Since $|B| = 2, \dim(\mathcal{V}) = 2$. (Hence, S does not span all of $\mathbb{R}^3$.) ■

[5] Theorem 4.17 is also true for infinite dimensional vector spaces, but the proof requires advanced topics in set theory that are beyond the scope of this book.

This method can also be adapted to vector spaces other than $\mathbb{R}^n$.

Example 5

Let $S = \{x^3 - 3x^2 + 1, 2x^2 + x, 2x^3 + 3x + 2, 4x - 5\} \subseteq \mathcal{P}_3$. We use the Independence Test Method to find a subset B of S that is a basis for $\mathcal{V} = \text{span}(S)$. Let $\mathbf{A}$ be the matrix whose columns are the analogous vectors in $\mathbb{R}^4$ for the given vectors in S. Then

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ -3 & 2 & 0 & 0 \\ 0 & 1 & 3 & 4 \\ 1 & 0 & 2 & -5 \end{bmatrix}, \quad \text{which reduces to} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Because we have nonzero pivots in the first, second, and fourth columns of $\mathbf{C}$, we choose $B = \{x^3 - 3x^2 + 1, 2x^2 + x, 4x - 5\}$. These are the first, second, and fourth vectors in S. Then B is the desired basis for $\mathcal{V}$.

The third vector in S is a linear combination of previous vectors in S. The first two entries of the third column of $\mathbf{C}$ give the coefficients of that linear combination; that is, $2x^3 + 3x + 2 = 2(x^3 - 3x^2 + 1) + 3(2x^2 + x)$.

The Simplified Span Method and the Independence Test Method for finding a basis are similar enough to cause confusion, so we contrast their various features in Table 4.2.

Shrinking an Infinite Spanning Set to a Basis

The Independence Test Method can sometimes be used successfully when the spanning set S is infinite.

Table 4.2 Contrasting the Simplified Span Method and Independence Test Method for finding a basis from a given spanning set S

Simplified Span Method	Independence Test Method
The vectors in S become the *rows* of a matrix.	The vectors in S become the *columns* of a matrix.
The basis created is *not* a subset of the spanning set S but contains vectors with a simpler form.	The basis created *is* a subset of the spanning set S.
The nonzero rows of the reduced row echelon form matrix are used as the basis vectors.	The pivot columns of the reduced row echelon form matrix are used to determine which vectors to select from S.

Example 6

Let $\mathcal{V}$ be the subspace of $\mathcal{M}_{22}$ consisting of all 2×2 symmetric matrices. Let S be the set of nonsingular matrices in $\mathcal{V}$, and let $\mathcal{W}$ = span(S) = span({nonsingular, symmetric 2×2 matrices}). We reduce S to a basis for $\mathcal{W}$ using the Independence Test Method, even though S is infinite. (We prove later that $\mathcal{W} = \mathcal{V}$, and so the basis we construct is actually a basis for $\mathcal{V}$.)

The strategy is to guess a *finite* subset Y of S that spans $\mathcal{W}$. We then use the Independence Test Method on Y to find the desired basis. We try to pick vectors for Y whose forms are as simple as possible to make computation easier. In this case, we choose the set of all nonsingular symmetric 2×2 matrices having only zeroes and ones as entries. That is,

$$Y = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \right\}.$$

Now, before continuing, we must ensure that span(Y) = $\mathcal{W}$. That is, we must show every nonsingular symmetric 2×2 matrix is in span(Y). In fact, we will show every symmetric 2×2 matrix is in span(Y) by finding real numbers w, x, y, and z so that

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = w\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + y\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + z\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Thus, we must prove that the system

$$\begin{cases} w + x \phantom{{}+y+z} = a \\ \phantom{w+{}} x + y + z = b \\ \phantom{w+{}} x + y + z = b \\ w \phantom{{}+x+y} + z = c \end{cases}$$

has solutions for w, x, y, and z in terms of a, b, and c. But $w = 0$, $x = a$, $y = b - a - c$, $z = c$ certainly satisfies the system. Hence, $\mathcal{V} \subseteq$ span(Y). Since span(Y) $\subseteq \mathcal{V}$, we have span(Y) = $\mathcal{V} = \mathcal{W}$.

We can now use the Independence Test Method on Y. We express the matrices in Y as corresponding vectors in $\mathbb{R}^4$ and create the matrix with these vectors as columns, as follows:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \text{which reduces to} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then, the desired basis is

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\},$$

the elements of Y corresponding to the pivot columns of C.

The method used in Example 6 is not guaranteed to work when the spanning set $\mathcal{S}$ has infinitely many elements because our choice for the finite set $\mathcal{Y}$ might not have the same span as $\mathcal{S}$. When this happens, the choice of a larger set $\mathcal{Y}$ may lead to success.

Finding a Basis from a Spanning Set by Inspection

When a spanning set $\mathcal{S}$ for a vector space $\mathcal{V}$ is given, it is sometimes easier to select a maximal linearly independent subset of $\mathcal{S}$ (and hence, a basis for $\mathcal{V}$) by process of elimination rather than row reduction. The idea behind the following method is to inspect each of the vectors in the given spanning set $\mathcal{S}$ in turn and eliminate any that are redundant; that is, any vectors in $\mathcal{S}$ that are linear combinations of previous vectors.

The formal technique presented in the following method resembles a proof by induction in that there is a "Base" Step followed by an "Inductive" Step that is repeated until the desired basis is found. The method stops when we run out of vectors to choose in the Inductive Step that are linearly independent of those previously chosen.[6]

Method for Finding a Basis from a Spanning Set by Inspection (Inspection Method)

Let $\mathcal{S}$ be a finite set of vectors spanning a vector space $\mathcal{V}$.

(1) **Base Step:** Choose $\mathbf{v}_1 \neq \mathbf{0}$ in $\mathcal{S}$.

Repeat the following step as many times as possible:

(2) **Inductive Step:** Assuming $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$ have already been chosen from $\mathcal{S}$, choose $\mathbf{v}_k \in \mathcal{S}$ such that $\mathbf{v}_k \notin \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$.

The final set constructed is a basis for $\mathcal{V}$.

The Inspection Method is useful when you can determine easily (without tedious computations) which vectors to choose next in the Inductive Step. Otherwise, you should apply the Independence Test Method.

Example 7

Let $\mathcal{S} = \{[0,0,0], [2,-8,12], [-1,4,-6], [7,2,2]\}$, a subset of $\mathbb{R}^3$. Let $\mathcal{V} = \text{span}(\mathcal{S})$, a subspace of $\mathbb{R}^3$. We use the Inspection Method to find a subset B of $\mathcal{S}$ that is a basis for $\mathcal{V}$.

The Base Step is to choose $\mathbf{v}_1$, a nonzero vector in $\mathcal{S}$. So, we skip over the first vector listed in $\mathcal{S}$, $[0,0,0]$ and let $\mathbf{v}_1 = [2,-8,12]$.

[6] We assume that $\mathcal{S}$ has at least one nonzero vector. Otherwise, $\mathcal{V}$ would be the trivial vector space. In this case, the desired basis for $\mathcal{V}$ is the empty set, { }.

Moving on to the Inductive Step, we look for $\mathbf{v}_2$ in S so that $\mathbf{v}_2 \notin \text{span}(\{\mathbf{v}_1\})$. Hence, $\mathbf{v}_2$ may not be a scalar multiple of $\mathbf{v}_1$. Therefore, we may not choose $[-1,4,-6]$ because $[-1,4,-6] = -\frac{1}{2}[2,-8,12]$. Instead, we choose $\mathbf{v}_2 = [7,2,2]$.

At this point, there are no more vectors in S for us to try, so the induction process must terminate here. Therefore, $B = \{\mathbf{v}_1, \mathbf{v}_2\} = \{[2,-8,12],[7,2,2]\}$ is the desired basis for $\mathcal{V}$. Notice that $\mathcal{V} = \text{span}(B)$ is not all of $\mathbb{R}^3$ because $\dim(\mathcal{V}) = 2 \neq \dim\left(\mathbb{R}^3\right)$. (You can verify, for example, that the vector $[1,0,0] \in \mathbb{R}^3$ cannot be expressed as a linear combination of the vectors in B and hence is not in $\mathcal{V} = \text{span}(B)$.)

Every Linearly Independent Set in a Finite Dimensional Vector Space Is Contained in Some Basis

Suppose that $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_k\}$ is a linearly independent set of vectors in a finite dimensional vector space $\mathcal{V}$. Because $\mathcal{V}$ is finite dimensional, it has a finite basis, say $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Consider the set $T \cup A$. Now, $T \cup A$ certainly spans $\mathcal{V}$ (since A alone spans $\mathcal{V}$). We can therefore apply the Independence Test Method to $T \cup A$ to produce a basis B for $\mathcal{V}$. If we order the vectors in $T \cup A$ so that all the vectors in T are listed first, then none of these vectors will be eliminated, since no vector in T is a linear combination of vectors listed earlier in T. In this manner we construct a basis B for $\mathcal{V}$ that contains T. We have just proved the following:

Theorem 4.18 Let T be a linearly independent subset of a finite dimensional vector space $\mathcal{V}$. Then $\mathcal{V}$ has a basis B with $T \subseteq B$.

Compare this result with Theorem 4.17.

We modify slightly the method outlined just before Theorem 4.18 to find a basis for a finite dimensional vector space containing a given linearly independent subset T.

Method for Finding a Basis by Enlarging a Linearly Independent Subset (Enlarging Method)
Suppose that $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_k\}$ is a linearly independent subset of a finite dimensional vector space $\mathcal{V}$.

Step 1: Find a finite spanning set $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ for $\mathcal{V}$.

Step 2: Form the ordered spanning set $S = \{\mathbf{t}_1, \ldots, \mathbf{t}_k, \mathbf{a}_1, \ldots, \mathbf{a}_n\}$ for $\mathcal{V}$.

Step 3: Use either the Independence Test Method or the Inspection Method on S to produce a subset B of S.

Then B is a basis for $\mathcal{V}$ containing T.

The basis produced by this method is easier to use if the additional vectors in the set A have a simple form. Ideally, we choose A to be the standard basis for $\mathcal{V}$.

Example 8

Consider the linearly independent subset $T = \{[2,0,4,-12],[0,-1,-3,9]\}$ of $\mathcal{V} = \mathbb{R}^4$. We use the Enlarging Method to find a basis for $\mathbb{R}^4$ that contains T.

Step 1: We choose A to be the standard basis $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4\}$ for $\mathbb{R}^4$.

Step 2: We create

$$S = \{[2,0,4,-12],[0,-1,-3,9],[1,0,0,0],[0,1,0,0],[0,0,1,0],[0,0,0,1]\}.$$

Step 3: The matrix

$$\begin{bmatrix} 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 4 & -3 & 0 & 0 & 1 & 0 \\ -12 & 9 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{reduces to} \quad \begin{bmatrix} 1 & 0 & 0 & -\frac{3}{4} & 0 & -\frac{1}{12} \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{3} \end{bmatrix}.$$

Since columns $1,2,3,$ and 5 have nonzero pivots, the Independence Test Method indicates that the set $B = \{[2,0,4,-12],[0,-1,-3,9],[1,0,0,0],[0,0,1,0]\}$ is a basis for $\mathbb{R}^4$ containing T.

In general, we can use the Enlarging Method only when we already know a finite spanning set to use for A. Otherwise, we can make an intelligent guess, just as we did when using the Independence Test Method on an infinite spanning set. However, we must then take care to verify that the resulting set actually spans the vector space.

New Vocabulary

Enlarging Method

Inspection Method

Highlights

- Every spanning set of a finite dimensional vector space $\mathcal{V}$ has a subset that is a basis for $\mathcal{V}$.
- Every linearly independent set of a finite dimensional vector space $\mathcal{V}$ can be enlarged to a basis for $\mathcal{V}$.
- The Simplified Span Method is useful for finding a basis (in simplified form) for the span of a given set of vectors (by row reducing the matrix whose rows are the given vectors).
- The Independence Test Method is useful for finding a *subset* of a given set of vectors that is a basis for the span of the vectors.

- The Inspection Method is useful for finding a subset of a spanning set that is a basis (by eliminating those vectors that are linear combinations of earlier vectors).
- The Enlarging Method is useful for enlarging a linearly independent set to a basis (for a finite dimensional vector space).

EXERCISES FOR SECTION 4.6

1. For each of the given subsets S of $\mathbb{R}^5$, find a basis for $\mathcal{V} = \text{span}(S)$ using the Simplified Span Method:

⋆(a) $S = \{[1,2,3,-1,0],[3,6,8,-2,0],[-1,-1,-3,1,1],[-2,-3,-5,1,1]\}$

(b) $S = \{[3,2,-1,0,1],[1,-1,0,3,1],[4,1,-1,3,2],[3,7,-2,-9,-1],$ $[-1,-4,1,6,1]\}$

(c) $S = \{[0,1,1,0,6],[2,-1,0,-2,1],[-1,2,1,1,2],[3,-2,0,-2,-3],$ $[1,1,1,-1,4],[2,-1,-1,1,3]\}$

⋆(d) $S = \{[1,1,1,1,1],[1,2,3,4,5],[0,1,2,3,4],[0,0,4,0,-1]\}$

⋆2. Adapt the Simplified Span Method to find a basis for the subspace of $\mathcal{P}_3$ spanned by $S = \{x^3 - 3x^2 + 2, 2x^3 - 7x^2 + x - 3, 4x^3 - 13x^2 + x + 5\}$.

⋆3. Adapt the Simplified Span Method to find a basis for the subspace of $\mathcal{M}_{32}$ spanned by

$$S = \left\{ \begin{bmatrix} 1 & 4 \\ 0 & -1 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 5 \\ 1 & -1 \\ 4 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 7 \\ -1 & -2 \\ 2 & -3 \end{bmatrix}, \begin{bmatrix} 3 & 6 \\ 2 & -1 \\ 6 & 12 \end{bmatrix} \right\}.$$

4. For each given subset S of $\mathbb{R}^3$, find a subset B of S that is a basis for $\mathcal{V} = \text{span}(S)$.

⋆(a) $S = \{[3,1,-2],[0,0,0],[6,2,-3]\}$

(b) $S = \{[4,7,1],[1,0,0],[6,7,1],[-4,0,0]\}$

⋆(c) $S = \{[1,3,-2],[2,1,4],[3,-6,18],[0,1,-1],[-2,1,-6]\}$

(d) $S = \{[1,4,-2],[-2,-8,4],[2,-8,5],[0,-7,2]\}$

⋆(e) $S = \{[3,-2,2],[1,2,-1],[3,-2,7],[-1,-10,6]\}$

(f) $S = \{[3,1,0],[2,-1,7],[0,0,0],[0,5,-21],[6,2,0],[1,5,7]\}$

(g) $S =$ the set of all 3-vectors whose second coordinate is zero

⋆(h) $S =$ the set of all 3-vectors whose second coordinate is -3 times its first coordinate plus its third coordinate

6.2 Diagonalization

5. For each given subset S of $\mathcal{P}_3$, find a subset B of S that is a basis for $\mathcal{V} = \text{span}(S)$.

★**(a)** $S = \{x^3 - 8x^2 + 1, 3x^3 - 2x^2 + x, 4x^3 + 2x - 10, x^3 - 20x^2 - x + 12, x^3 + 24x^2 + 2x - 13, x^3 + 14x^2 - 7x + 18\}$

(b) $S = \{-2x^3 + x + 2, 3x^3 - x^2 + 4x + 6, 8x^3 + x^2 + 6x + 10, -4x^3 - 3x^2 + 3x + 4, -3x^3 - 4x^2 + 8x + 12\}$

★**(c)** $S =$ the set of all polynomials in $\mathcal{P}_3$ with a zero constant term

(d) $S = \mathcal{P}_2$

★**(e)** $S =$ the set of all polynomials in $\mathcal{P}_3$ with the coefficient of the x^2 term equal to the coefficient of the x^3 term

(f) $S =$ the set of all polynomials in $\mathcal{P}_3$ with the coefficient of the x^3 term equal to 8

6. For each given subset S of $\mathcal{M}_{33}$, find a subset B of S that is a basis for $\mathcal{V} = \text{span}(S)$.

★**(a)** $S = \{\mathbf{A} \in \mathcal{M}_{33} |$ each a_{ij} is either 0 or 1$\}$

(b) $S = \{\mathbf{A} \in \mathcal{M}_{33} |$ each a_{ij} is either 1 or $-1\}$

★**(c)** $S =$ the set of all symmetric 3×3 matrices

(d) $S =$ the set of all nonsingular 3×3 matrices

7. Enlarge each of the following linearly independent subsets T of $\mathbb{R}^5$ to a basis B for $\mathbb{R}^5$ containing T:

★**(a)** $T = \{[1, -3, 0, 1, 4], [2, 2, 1, -3, 1]\}$

(b) $T = \{[1, 1, 1, 1, 1], [0, 1, 1, 1, 1], [0, 0, 1, 1, 1]\}$

★**(c)** $T = \{[1, 0, -1, 0, 0], [0, 1, -1, 1, 0], [2, 3, -8, -1, 0]\}$

8. Enlarge each of the following linearly independent subsets T of $\mathcal{P}_4$ to a basis B for $\mathcal{P}_4$ that contains T:

★**(a)** $T = \{x^3 - x^2, x^4 - 3x^3 + 5x^2 - x\}$

(b) $T = \{3x - 2, x^3 - 6x + 4\}$

★**(c)** $T = \{x^4 - x^3 + x^2 - x + 1, x^3 - x^2 + x - 1, x^2 - x + 1\}$

9. Enlarge each of the following linearly independent subsets T of $\mathcal{M}_{32}$ to a basis B for $\mathcal{M}_{32}$ that contains T:

★**(a)** $T = \left\{ \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -1 & 1 \end{bmatrix} \right\}$

(b) $T = \left\{ \begin{bmatrix} 0 & -2 \\ 1 & 0 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & -3 \\ 0 & 1 \\ 3 & -6 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ -4 & 8 \end{bmatrix} \right\}$

★(c) $T = \left\{ \begin{bmatrix} 3 & 0 \\ -1 & 7 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 3 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 3 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 6 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \right\}$

★10. Find a basis for the vector space $\mathcal{U}_4$ consisting of all 4×4 upper triangular matrices.

11. In each case, find the dimension of $\mathcal{V}$ by using an appropriate method to create a basis.

(a) $\mathcal{V} = \text{span}(\{[5,2,1,0,-1],[3,0,1,1,0],[0,0,0,0,0],[-2,4,-2,-4,-2],$ $[0,12,-4,-10,-6],[-6,0,-2,-2,0]\})$, a subspace of $\mathbb{R}^5$

★(b) $\mathcal{V} = \{\mathbf{A} \in \mathcal{M}_{33} \mid \text{trace}(\mathbf{A}) = 0\}$, a subspace of $\mathcal{M}_{33}$ (Recall that the trace of a matrix is the sum of the terms on the main diagonal.)

(c) $\mathcal{V} = \text{span}(\{x^4 - x^3 + 2x^2, 2x^4 + x - 5, 2x^3 - 4x^2 + x - 4, 6, x^2 - 1\})$

★(d) $\mathcal{V} = \{\mathbf{p} \in \mathcal{P}_6 \mid \mathbf{p} = ax^6 - bx^5 + ax^4 - cx^3 + (a+b+c)x^2 - (a-c)x + (3a - 2b + 16c)$, for real numbers a, b, and $c\}$

12. **(a)** Show that each of these subspaces of $\mathcal{M}_{nn}$ has dimension $(n^2 + n)/2$.

(i) The set of upper triangular $n \times n$ matrices

(ii) The set of lower triangular $n \times n$ matrices

(iii) The set of symmetric $n \times n$ matrices

★(b) What is the dimension of the set of skew-symmetric $n \times n$ matrices?

13. Let $\mathbf{A}$ be an $m \times n$ matrix.

(a) Prove that $S_{\mathbf{A}} = \{\mathbf{X} \in \mathbb{R}^n \mid \mathbf{AX} = \mathbf{0}\}$, the solution set of the homogeneous system $\mathbf{AX} = \mathbf{0}$, is a subspace of $\mathbb{R}^n$.

(b) Prove that $\dim(S_{\mathbf{A}}) + \text{rank}(\mathbf{A}) = n$. (Hint: First consider the case where $\mathbf{A}$ is in reduced row echelon form.)

▶14. Prove Theorem 4.17. This proof should be similar to the part of the proof for Theorem 4.16 outlined in parts (a), (b), and (c) of Exercise 22 in Section 4.5. However, change the definition of the set A in that exercise so that each set T is a subset of S rather than of $\mathcal{W}$.

15. Let $\mathcal{W}$ be a subspace of a finite dimensional vector space $\mathcal{V}$.

(a) Show that $\mathcal{V}$ has some basis B with a subset B' that is a basis for $\mathcal{W}$.

★(b) If B is any given basis for $\mathcal{V}$, must some subset B' of B be a basis for $\mathcal{W}$? Prove that your answer is correct.

★(c) If B is any given basis for $\mathcal{V}$ and $B' \subseteq B$, is there necessarily a subspace $\mathcal{Y}$ of $\mathcal{V}$ such that B' is a basis for $\mathcal{Y}$? Why or why not?

16. Let $\mathcal{V}$ be a finite dimensional vector space, and let $\mathcal{W}$ be a subspace of $\mathcal{V}$.

(a) Prove that $\mathcal{V}$ has a subspace $\mathcal{W}'$ such that every vector in $\mathcal{V}$ can be uniquely expressed as a sum of a vector in $\mathcal{W}$ and a vector in $\mathcal{W}'$. (In other words, show that there is a subspace $\mathcal{W}'$ so that, for every $\mathbf{v}$ in $\mathcal{V}$, there are unique vectors $\mathbf{w} \in \mathcal{W}$ and $\mathbf{w}' \in \mathcal{W}'$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}'$.)

⋆**(b)** Give an example of a subspace $\mathcal{W}$ of some finite dimensional vector space $\mathcal{V}$ for which the subspace $\mathcal{W}'$ from part (a) is not unique.

17. **(a)** Let S be a finite subset of $\mathbb{R}^n$. Prove that the Simplified Span Method applied to S produces the standard basis for $\mathbb{R}^n$ if and only if $\text{span}(S) = \mathbb{R}^n$.

(b) Let $B \subseteq \mathbb{R}^n$ with $|B| = n$, and let $\mathbf{A}$ be the $n \times n$ matrix whose rows are the vectors in B. Prove that B is a basis for $\mathbb{R}^n$ if and only if $|\mathbf{A}| \neq 0$.

18. Let $\mathbf{A}$ be an $m \times n$ matrix and let S be the set of vectors consisting of the rows of $\mathbf{A}$.

(a) Use the Simplified Span Method to show that $\dim(\text{span}(S)) = \text{rank}(\mathbf{A})$.

(b) Use the Independence Test Method to prove that $\dim(\text{span}(S)) = \text{rank}(\mathbf{A}^T)$.

(c) Use parts (a) and (b) to prove that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$. (We will state this formally as Corollary 5.11 in Section 5.3.)

19. Let $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ be any real numbers, with $n > 2$. Consider the $n \times n$ matrix $\mathbf{A}$ whose (i,j) term is $a_{ij} = \sin(\alpha_i + \beta_j)$. Prove that $|\mathbf{A}| = 0$. (Hint: Consider $\mathbf{x}_1 = [\sin\beta_1, \sin\beta_2, \ldots, \sin\beta_n]$, $\mathbf{x}_2 = [\cos\beta_1, \cos\beta_2, \ldots, \cos\beta_n]$. Show that the row space of $\mathbf{A} \subseteq \text{span}(\{\mathbf{x}_1, \mathbf{x}_2\})$, and hence, dim(row space of $\mathbf{A}$) $< n$.)

⋆**20.** True or False:

(a) Given any spanning set S for a finite dimensional vector space $\mathcal{V}$, there is some $B \subseteq S$ that is a basis for $\mathcal{V}$.

(b) Given any linearly independent set T in a finite dimensional vector space $\mathcal{V}$, there is a basis B for $\mathcal{V}$ containing T.

(c) If S is a finite spanning set for $\mathbb{R}^n$, then the Simplified Span Method must produce a subset of S that is a basis for $\mathbb{R}^n$.

(d) If S is a finite spanning set for $\mathbb{R}^n$, then the Independence Test Method produces a subset of S that is a basis for $\mathbb{R}^n$.

(e) If S is a finite spanning set for $\mathbb{R}^n$, then the Inspection Method produces a subset of S that is a basis for $\mathbb{R}^n$.

(f) If T is a linearly independent set in $\mathbb{R}^n$, then the Enlarging Method must produce a subset of T that is a basis for $\mathbb{R}^n$.

(g) Before row reduction, the Simplified Span Method places the vectors of a given spanning set S as columns in a matrix, while the Independence Test Method places the vectors of S as rows.

4.7 COORDINATIZATION

If B is a basis for a vector space $\mathcal{V}$, then we know every vector in $\mathcal{V}$ has a unique expression as a linear combination of the vectors in B. For example, the vector $[a_1,\ldots,a_n]$ in $\mathbb{R}^n$ is written as a linear combination of the standard basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for $\mathbb{R}^n$ in a natural and unique way as $a_1\mathbf{e}_1+\cdots+a_n\mathbf{e}_n$. Dealing with the standard basis in $\mathbb{R}^n$ is easy because the coefficients in the linear combination are the same as the coordinates of the vector. However, this is not necessarily true for other bases.

In this section, we develop a process, called coordinatization, for representing any vector in a finite dimensional vector space in terms of its coefficients with respect to a given basis. We also determine how the coordinatization changes whenever we switch bases.

Coordinates with Respect to an Ordered Basis

Definition An **ordered basis** for a vector space $\mathcal{V}$ is an ordered n-tuple of vectors $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ such that the set $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ is a basis for $\mathcal{V}$.

In an ordered basis, the elements are written in a specific order. Thus, $(\mathbf{i},\mathbf{j},\mathbf{k})$ and $(\mathbf{j},\mathbf{i},\mathbf{k})$ are different ordered bases for $\mathbb{R}^3$.

By Theorem 4.9, if $B=(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n)$ is an ordered basis for $\mathcal{V}$, then for every vector $\mathbf{w}\in\mathcal{V}$, there are unique scalars $a_1,a_2,\ldots,a_n$ such that $\mathbf{w}=a_1\mathbf{v}_1+a_2\mathbf{v}_2+\cdots+a_n\mathbf{v}_n$. We use these scalars $a_1,a_2,\ldots,a_n$ to **coordinatize** the vector $\mathbf{w}$ as follows:

Definition Let $B=(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n)$ be an ordered basis for a vector space $\mathcal{V}$. Suppose that $\mathbf{w}=a_1\mathbf{v}_1+a_2\mathbf{v}_2+\cdots+a_n\mathbf{v}_n\in\mathcal{V}$. Then $[\mathbf{w}]_B$, the **coordinatization of w with respect to** B, is the n-vector $[a_1,a_2,\ldots,a_n]$.

The vector $[\mathbf{w}]_B=[a_1,a_2,\ldots,a_n]$ is frequently referred to as "**w expressed in** B**-coordinates**." When useful, we will express $[\mathbf{w}]_B$ as a column vector.

Example 1

Let $B=([4,2],[1,3])$ be an ordered basis for $\mathbb{R}^2$. Notice that $[4,2]=1[4,2]+0[1,3]$, so $[4,2]_B=[1,0]$. Similarly, $[1,3]_B=[0,1]$. From a geometric viewpoint, converting to B-coordinates in $\mathbb{R}^2$ results in a new coordinate system in $\mathbb{R}^2$ with $[4,2]$ and $[1,3]$ as its "unit" vectors. The new coordinate grid consists of parallelograms whose sides are the vectors in B, as shown in Figure 4.6. For example, $[11,13]$ equals $[2,3]$ when expressed in B-coordinates because $[11,13]=2[4,2]+3[1,3]$. In other words, $[11,13]_B=[2,3]$.

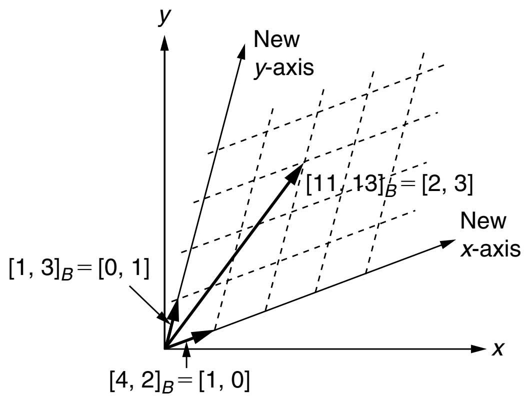

FIGURE 4.6

A B-coordinate grid in $\mathbb{R}^2$: picturing $[11,13]$ in B-coordinates

Example 2

Let $B = (x^3, x^2, x, 1)$, an ordered basis for $\mathcal{P}_3$. Then $[6x^3 - 2x + 18]_B = [6, 0, -2, 18]$, and $[4 - 3x + 9x^2 - 7x^3]_B = [-7, 9, -3, 4]$. Notice also that $[x^3]_B = [1,0,0,0], [x^2]_B = [0,1,0,0]$, $[x]_B = [0,0,1,0]$, and $[1]_B = [0,0,0,1]$.

As part of Example 2, we saw an illustration of the general principle that if $B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, then every vector in B itself has a simple coordinatization. In particular, $[\mathbf{v}_i]_B = \mathbf{e}_i$. You are asked to prove this in Exercise 6.

Using Row Reduction to Coordinatize a Vector

Example 3

Consider the subspace $\mathcal{V}$ of $\mathbb{R}^5$ spanned by the ordered basis

$$C = \big([-4,5,-1,0,-1],[1,-3,2,2,5],[1,-2,1,1,3]\big).$$

Notice that the vectors in $\mathcal{V}$ can be put into C-coordinates by solving an appropriate system. For example, to find $[-23, 30, -7, -1, -7]_C$, we solve the equation

$$[-23,30,-7,-1,-7] = a[-4,5,-1,0,-1] + b[1,-3,2,2,5] + c[1,-2,1,1,3].$$

The equivalent system is

$$\begin{cases} -4a + b + c = -23 \\ 5a - 3b - 2c = 30 \\ -a + 2b + c = -7 \\ 2b + c = -1 \\ -a + 5b + 3c = -7 \end{cases}.$$

To solve this system, we row reduce

$$\left[\begin{array}{rrr|r} -4 & 1 & 1 & -23 \\ 5 & -3 & -2 & 30 \\ -1 & 2 & 1 & -7 \\ 0 & 2 & 1 & -1 \\ -1 & 5 & 3 & -7 \end{array}\right] \quad \text{to obtain} \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

Hence, the (unique) solution for the system is $a = 6, b = -2, c = 3$, and we see that $[-23, 30, -7, -1, -7]_C = [6, -2, 3]$.

On the other hand, vectors in $\mathbb{R}^5$ that are not in span(C) cannot be expressed in C-coordinates. For example, the vector $[1, 2, 3, 4, 5]$ is not in $\mathcal{V} = \text{span}(C)$. To see this, consider the system

$$\begin{cases} -4a + b + c = 1 \\ 5a - 3b - 2c = 2 \\ {-a} + 2b + c = 3 \\ \phantom{-4a+{}} 2b + c = 4 \\ {-a} + 5b + 3c = 5 \end{cases}.$$

We solve this system by row reducing

$$\left[\begin{array}{rrr|r} -4 & 1 & 1 & 1 \\ 5 & -3 & -2 & 2 \\ -1 & 2 & 1 & 3 \\ 0 & 2 & 1 & 4 \\ -1 & 5 & 3 & 5 \end{array}\right] \quad \text{to obtain} \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

This result tells us that the system has no solutions, implying that the vector $[1, 2, 3, 4, 5]$ is not in span(S).

$\blacksquare$

Notice in Example 3 that the coordinatized vector $[6, -2, 3]$ is more "compact" than the original vector $[-23, 30, -7, -1, -7]$ but still contains the same essential information.

As we saw in Example 3, finding the coordinates of a vector with respect to an ordered basis typically amounts to solving a system of linear equations, which is frequently done using row reduction. The computations we did in Example 3 motivate the following method, which works in general. Although it applies to subspaces of $\mathbb{R}^n$, we can adapt it to other finite dimensional vector spaces, such as $\mathcal{P}_n$ and $\mathcal{M}_{mn}$, as with other techniques we have examined. We handle these other vector spaces "informally" in this chapter, but we will treat them more formally in Section 5.5.

Method for Coordinatizing a Vector with Respect to a Finite Ordered Basis (Coordinatization Method)

Let $\mathcal{V}$ be a nontrivial subspace of $\mathbb{R}^n$, let $B = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$ be an ordered basis for $\mathcal{V}$, and let $\mathbf{v} \in \mathbb{R}^n$. To calculate $[\mathbf{v}]_B$, if possible, perform the following steps:

Step 1: Form an augmented matrix $[\mathbf{A} \mid \mathbf{v}]$ by using the vectors in B as the *columns* of $\mathbf{A}$, in *order*, and using $\mathbf{v}$ as a column on the right.

Step 2: Row reduce $[\mathbf{A}\,|\,\mathbf{v}]$ to obtain the reduced row echelon form $[\mathbf{C}\,|\,\mathbf{w}]$.

Step 3: If there is a row of $[\mathbf{C}\,|\,\mathbf{w}]$ that contains all zeroes on the left and has a nonzero entry on the right, then $\mathbf{v} \notin \text{span}(B) = \mathcal{V}$, and coordinatization is not possible. Stop.

Step 4: Otherwise, $\mathbf{v} \in \text{span}(B) = \mathcal{V}$. Eliminate all rows consisting entirely of zeroes in $[\mathbf{C}\,|\,\mathbf{w}]$ to obtain $[\mathbf{I}_k\,|\,\mathbf{y}]$. Then, $[\mathbf{v}]_B = \mathbf{y}$, the last column of $[\mathbf{I}_k\,|\,\mathbf{y}]$.

Example 4

Let $\mathcal{V}$ be the subspace of $\mathbb{R}^3$ spanned by the ordered basis

$$B = ([2,-1,3],[3,2,1]).$$

We use the Coordinatization Method to find $[\mathbf{v}]_B$, where $\mathbf{v} = [5,-6,11]$. To do this, we set up the augmented matrix

$$\left[\begin{array}{rr|r} 2 & 3 & 5 \\ -1 & 2 & -6 \\ 3 & 1 & 11 \end{array}\right], \quad \text{which row reduces to} \quad \left[\begin{array}{rr|r} 1 & 0 & 4 \\ 0 & 1 & -1 \\ \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

Ignoring the bottom row of zeroes, we discover $[\mathbf{v}]_B = [4,-1]$.

Similarly, applying the Coordinatization Method to the vector $[1,2,3]$, we see that

$$\left[\begin{array}{rr|r} 2 & 3 & 1 \\ -1 & 2 & 2 \\ 3 & 1 & 3 \end{array}\right] \quad \text{row reduces to} \quad \left[\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

From the third row, we see that coordinatization of $[1,2,3]$ with respect to B is not possible by Step 3 of the Coordinatization Method. ■

Fundamental Properties of Coordinatization

The following theorem shows that the coordinatization of a vector behaves in a manner similar to the original vector with respect to addition and scalar multiplication:

Theorem 4.19 Let $B = (\mathbf{v}_1,\ldots,\mathbf{v}_n)$ be an ordered basis for a vector space $\mathcal{V}$. Suppose $\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathcal{V}$ and $a_1,\ldots,a_k$ are scalars. Then

(1) $[\mathbf{w}_1 + \mathbf{w}_2]_B = [\mathbf{w}_1]_B + [\mathbf{w}_2]_B$

(2) $[a_1\mathbf{w}_1]_B = a_1[\mathbf{w}_1]_B$

(3) $[a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k]_B = a_1[\mathbf{w}_1]_B + a_2[\mathbf{w}_2]_B + \cdots + a_k[\mathbf{w}_k]_B$

Figure 4.7 illustrates part (1) of this theorem. Moving along either path from the upper left to the lower right in the diagram produces the same answer. (Such a picture is called a **commutative diagram**.)

Part (3) asserts that to put a linear combination of vectors in $\mathcal{V}$ into B-coordinates, we can first find the B-coordinates of each vector individually and then calculate the analogous linear combination in $\mathbb{R}^n$. The proof of Theorem 4.19 is left for you to do in Exercise 13.

Example 5

Recall the subspace $\mathcal{V}$ of $\mathbb{R}^5$ from Example 3 spanned by the ordered basis

$$C = \big([-4,5,-1,0,-1],[1,-3,2,2,5],[1,-2,1,1,3]\big).$$

Consider the vectors $\mathbf{x} = [1,0,-1,0,4], \mathbf{y} = [0,1,-1,0,3], \mathbf{z} = [0,0,0,1,5]$. Applying the Coordinatization Method to $\mathbf{x}$, we find that the augmented matrix

$$\left[\begin{array}{rrr|r} -4 & 1 & 1 & 1 \\ 5 & -3 & -2 & 0 \\ -1 & 2 & 1 & -1 \\ 0 & 2 & 1 & 0 \\ -1 & 5 & 3 & 4 \end{array}\right] \quad \text{row reduces to} \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 10 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

Ignoring the last two rows of zeroes, we obtain $[\mathbf{x}]_C = [1,-5,10]$. In a similar manner we can calculate $[\mathbf{y}]_C = [1,-4,8]$ and $[\mathbf{z}]_C = [1,-3,7]$.

Using Theorem 4.19, it is now a simple matter to find the coordinatization of any linear combination of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. For example, consider the vector $2\mathbf{x} - 7\mathbf{y} + 3\mathbf{z}$, which is easily computed to be $[2,-7,5,3,2]$. Theorem 4.19 asserts that $[2\mathbf{x} - 7\mathbf{y} + 3\mathbf{z}]_C = 2[\mathbf{x}]_C - 7[\mathbf{y}]_C + 3[\mathbf{z}]_C = 2[1,-5,10] - 7[1,-4,8] + 3[1,-3,7] = [-2,9,-15]$. This result is easily checked by noting that $-2[-4,5,-1,0,-1] + 9[1,-3,2,2,5] - 15[1,-2,1,1,3]$ really does equal $[2,-7,5,3,2]$.

$$\begin{tikzcd}
\overset{\mathbf{w}_1,\,\mathbf{w}_2}{\mathcal{V}} \arrow[r, "+"] \arrow[d, "\text{Coordinatization}"'] & \overset{\mathbf{w}_1+\mathbf{w}_2}{\mathcal{V}} \arrow[d, "\text{Coordinatization}"] \\
\underset{[\mathbf{w}_1]_B,\,[\mathbf{w}_2]_B}{\mathbb{R}^n} \arrow[r, "+"] & \underset{[\mathbf{w}_1]_B+[\mathbf{w}_2]_B}{\mathbb{R}^n}
\end{tikzcd}$$

FIGURE 4.7

Commutative diagram involving addition and coordinatization of vectors

The Transition Matrix for Change of Coordinates

Our next goal is to determine how the coordinates of a vector change when we convert from one ordered basis to another.

Definition Suppose that $\mathcal{V}$ is a nontrivial n-dimensional vector space with ordered bases B and C. Let $\mathbf{P}$ be the $n \times n$ matrix whose ith column, for $1 \le i \le n$, equals $[\mathbf{b}_i]_C$, where $\mathbf{b}_i$ is the ith basis vector in B. Then $\mathbf{P}$ is called the **transition matrix from B-coordinates to C-coordinates**.

We often refer to the matrix $\mathbf{P}$ in this definition as the "**transition matrix from B to C**."

Example 6

Recall from Example 5 the subspace $\mathcal{V}$ of $\mathbb{R}^5$ that is spanned by the ordered basis $C = ([-4,5,-1,0,-1],[1,-3,2,2,5],[1,-2,1,1,3])$. Using the Simplified Span Method on the vectors in C produces the vectors $\mathbf{x} = [1,0,-1,0,4], \mathbf{y} = [0,1,-1,0,3]$, and $\mathbf{z} = [0,0,0,1,5]$ from Example 5. Thus $B = (\mathbf{x},\mathbf{y},\mathbf{z})$ is also an ordered basis for $\mathcal{V}$. To find the transition matrix from B to C we must solve for the C-coordinates of each vector in B. In Example 5, we used the Coordinatization Method on each of $\mathbf{x},\mathbf{y}$, and $\mathbf{z}$ in turn. However, we could have obtained the same result by applying the Coordinatization Method to $\mathbf{x},\mathbf{y}$, and $\mathbf{z}$ simultaneously — that is, by row reducing the augmented matrix

$$\left[\begin{array}{rrr|rrr} -4 & 1 & 1 & 1 & 0 & 0 \\ 5 & -3 & -2 & 0 & 1 & 0 \\ -1 & 2 & 1 & -1 & -1 & 0 \\ 0 & 2 & 1 & 0 & 0 & 1 \\ -1 & 5 & 3 & 4 & 3 & 5 \end{array}\right] \quad \text{to obtain} \quad \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -5 & -4 & -3 \\ 0 & 0 & 1 & 10 & 8 & 7 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

This gives $[\mathbf{x}]_C = [1,-5,10], [\mathbf{y}]_C = [1,-4,8]$, and $[\mathbf{z}]_C = [1,-3,7]$ (as we saw earlier). These vectors form the columns of the transition matrix from B to C, namely,

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 1 \\ -5 & -4 & -3 \\ 10 & 8 & 7 \end{bmatrix}.$$

Example 6 illustrates that solving for the columns of the transition matrix can be accomplished efficiently by performing a single row reduction using an augmented matrix with several columns to the right of the augmentation bar. Hence, we have the following:

Method for Calculating a Transition Matrix (Transition Matrix Method)
To find the transition matrix $\mathbf{P}$ from B to C where B and C are ordered bases for a nontrivial k-dimensional subspace of $\mathbb{R}^n$, use row reduction on

$$\left[\begin{array}{cccc|cccc} \text{1st} & \text{2nd} & & k\text{th} & \text{1st} & \text{2nd} & & k\text{th} \\ \text{vector} & \text{vector} & \cdots & \text{vector} & \text{vector} & \text{vector} & \cdots & \text{vector} \\ \text{in} & \text{in} & & \text{in} & \text{in} & \text{in} & & \text{in} \\ C & C & & C & B & B & & B \end{array}\right]$$

to produce $\left[\begin{array}{c|c} \mathbf{I}_k & \mathbf{P} \\ \hline \text{rows of} & \text{zeroes} \end{array}\right]$.

In Exercise 8, you are asked to show that, in the special cases where either B or C is the standard basis in $\mathbb{R}^n$, there are simple expressions for the transition matrix from B to C.

Example 7

Consider the following ordered bases for $\mathcal{U}_2$:

$$B = \left(\begin{bmatrix} 7 & 3 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\right) \quad \text{and} \quad C = \left(\begin{bmatrix} 22 & 7 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 12 & 4 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 33 & 12 \\ 0 & 2 \end{bmatrix}\right).$$

Expressing the matrices in B and C as column vectors, we use the Transition Matrix Method to find the transition matrix from B to C by row reducing

$$\left[\begin{array}{ccc|ccc} 22 & 12 & 33 & 7 & 1 & 1 \\ 7 & 4 & 12 & 3 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 2 & 0 & -1 & 1 \end{array}\right] \quad \text{to obtain} \quad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & 1 \\ 0 & 1 & 0 & -4 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{array}\right].$$

Ignoring the final row of zeroes, we see that the transition matrix from B to C is given by

$$\mathbf{P} = \begin{bmatrix} 1 & -2 & 1 \\ -4 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

Change of Coordinates Using the Transition Matrix

The next theorem shows that the transition matrix can be used to change the coordinatization of a vector $\mathbf{v}$ from one ordered basis B to another ordered basis C. That is, if $[\mathbf{v}]_B$ is known, then $[\mathbf{v}]_C$ can be found by using the transition matrix from B to C.

Theorem 4.20 Suppose that B and C are ordered bases for a nontrivial n-dimensional vector space $\mathcal{V}$, and let $\mathbf{P}$ be an $n \times n$ matrix. Then $\mathbf{P}$ is the transition matrix from B to C if and only if for every $\mathbf{v} \in \mathcal{V}$, $\mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C$.

Proof. Let B and C be ordered bases for a vector space $\mathcal{V}$, with $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$.

First, suppose $\mathbf{P}$ is the transition matrix from B to C. Let $\mathbf{v} \in \mathcal{V}$. We want to show $\mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C$. Suppose $[\mathbf{v}]_B = [a_1, \ldots, a_n]$. Then $\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$. Hence,

$$\mathbf{P}[\mathbf{v}]_B = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$= a_1 \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{bmatrix} + a_2 \begin{bmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{n2} \end{bmatrix} + \cdots + a_n \begin{bmatrix} p_{1n} \\ p_{2n} \\ \vdots \\ p_{nn} \end{bmatrix}.$$

However, $\mathbf{P}$ is the transition matrix from B to C, so the ith column of $\mathbf{P}$ equals $[\mathbf{b}_i]_C$. Therefore,

$$\begin{aligned} \mathbf{P}[\mathbf{v}]_B &= a_1[\mathbf{b}_1]_C + a_2[\mathbf{b}_2]_C + \cdots + a_n[\mathbf{b}_n]_C \\ &= [a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n]_C \qquad \text{by Theorem 4.19} \\ &= [\mathbf{v}]_C. \end{aligned}$$

Conversely, suppose that $\mathbf{P}$ is an $n \times n$ matrix and that $\mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C$ for every $\mathbf{v} \in \mathcal{V}$. We show that $\mathbf{P}$ is the transition matrix from B to C. By definition, it is enough to show that the ith column of $\mathbf{P}$ is equal to $[\mathbf{b}_i]_C$. Since $\mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C$, for all $\mathbf{v} \in \mathcal{V}$, let $\mathbf{v} = \mathbf{b}_i$. Then since $[\mathbf{v}]_B = \mathbf{e}_i$, we have $\mathbf{P}[\mathbf{v}]_B = \mathbf{P}\mathbf{e}_i = [\mathbf{b}_i]_C$. But $\mathbf{P}\mathbf{e}_i = i$th column of $\mathbf{P}$, which completes the proof. □

Example 8

Recall the ordered bases for $\mathcal{U}_2$ from Example 7:

$$B = \left(\begin{bmatrix} 7 & 3 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right) \quad \text{and} \quad C = \left(\begin{bmatrix} 22 & 7 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 12 & 4 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 33 & 12 \\ 0 & 2 \end{bmatrix} \right).$$

In that example, we found that the transition matrix $\mathbf{P}$ from B to C is

$$\mathbf{P} = \begin{bmatrix} 1 & -2 & 1 \\ -4 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

This gives a quick way of changing the coordinatization of any vector in $\mathcal{U}_2$ from B-coordinates to C-coordinates. For example, let $\mathbf{v} = \begin{bmatrix} 25 & 24 \\ 0 & -9 \end{bmatrix}$. Since

$$\begin{bmatrix} 25 & 24 \\ 0 & -9 \end{bmatrix} = 4\begin{bmatrix} 7 & 3 \\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} - 6\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix},$$

we know that

$$[\mathbf{v}]_B = \begin{bmatrix} 4 \\ 3 \\ -6 \end{bmatrix}. \quad \text{But then,} \quad \mathbf{P}[\mathbf{v}]_B = \begin{bmatrix} -8 \\ -19 \\ 13 \end{bmatrix},$$

and so $[\mathbf{v}]_C = [-8, -19, 13]$ by Theorem 4.20. We can easily verify this by checking that

$$\begin{bmatrix} 25 & 24 \\ 0 & -9 \end{bmatrix} = -8\begin{bmatrix} 22 & 7 \\ 0 & 2 \end{bmatrix} - 19\begin{bmatrix} 12 & 4 \\ 0 & 1 \end{bmatrix} + 13\begin{bmatrix} 33 & 12 \\ 0 & 2 \end{bmatrix}.$$

Algebra of the Transition Matrix

The next theorem shows that the cumulative effect of two transitions between bases is represented by the product of the transition matrices in *reverse* order.

Theorem 4.21 Suppose that B, C, and D are ordered bases for a nontrivial finite dimensional vector space $\mathcal{V}$. Let $\mathbf{P}$ be the transition matrix from B to C, and let $\mathbf{Q}$ be the transition matrix from C to D. Then $\mathbf{QP}$ is the transition matrix from B to D.

The proof of this theorem is left as Exercise 14.

Example 9

Consider the ordered bases B and C for $\mathcal{P}_2$ given by

$$B = (-x^2 + 4x + 2,\ 2x^2 - x - 1,\ -x^2 + 2x + 1) \quad \text{and}$$

$$C = (x^2 - 2x - 3,\ 2x^2 - 1,\ x^2 + x + 1).$$

Also consider the standard basis $S = (x^2, x, 1)$ for $\mathcal{P}_2$.

Now, row reducing

$$\left[\begin{array}{rrr|rrr} 1 & 2 & 1 & -1 & 2 & -1 \\ -2 & 0 & 1 & 4 & -1 & 2 \\ -3 & -1 & 1 & 2 & -1 & 1 \end{array}\right] \quad \text{to obtain} \quad \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -9 & 3 & -5 \\ 0 & 1 & 0 & 11 & -3 & 6 \\ 0 & 0 & 1 & -14 & 5 & -8 \end{array}\right],$$

we see that the transition matrix from B to C is

$$\mathbf{P} = \begin{bmatrix} -9 & 3 & -5 \\ 11 & -3 & 6 \\ -14 & 5 & -8 \end{bmatrix}.$$

Because it is simple to express each vector in C in S-coordinates, we can quickly calculate that the transition matrix from C to S is

$$\mathbf{Q} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & 0 & 1 \\ -3 & -1 & 1 \end{bmatrix}.$$

Then, by Theorem 4.21, the product

$$\mathbf{QP} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & 0 & 1 \\ -3 & -1 & 1 \end{bmatrix} \begin{bmatrix} -9 & 3 & -5 \\ 11 & -3 & 6 \\ -14 & 5 & -8 \end{bmatrix} = \begin{bmatrix} -1 & 2 & -1 \\ 4 & -1 & 2 \\ 2 & -1 & 1 \end{bmatrix}$$

is the transition matrix from B to S. This matrix is correct, since the columns of $\mathbf{QP}$ are, in fact, the vectors of B expressed in S-coordinates. ■

The next theorem shows how to reverse a transition from one basis to another. The proof of this theorem is left as Exercise 15.

Theorem 4.22 Let B and C be ordered bases for a nontrivial finite dimensional vector space $\mathcal{V}$, and let $\mathbf{P}$ be the transition matrix from B to C. Then $\mathbf{P}$ is nonsingular, and $\mathbf{P}^{-1}$ is the transition matrix from C to B.

Let us return to the situation in Example 9 and use the inverses of the transition matrices to find the B-coordinates of a polynomial in $\mathcal{P}_2$.

Example 10

Consider again the bases B, C, and S in Example 9 and the transition matrices $\mathbf{P}$ from B to C and $\mathbf{Q}$ from C to S. From Theorem 4.22, the transition matrices from C to B and from S to C, respectively, are

$$\mathbf{P}^{-1} = \begin{bmatrix} -6 & -1 & 3 \\ 4 & 2 & -1 \\ 13 & 3 & -6 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{-1} = \begin{bmatrix} 1 & -3 & 2 \\ -1 & 4 & -3 \\ 2 & -5 & 4 \end{bmatrix}.$$

Now,

$$[\mathbf{v}]_B = \mathbf{P}^{-1}[\mathbf{v}]_C = \mathbf{P}^{-1}\left(\mathbf{Q}^{-1}[\mathbf{v}]_S\right) = \left(\mathbf{P}^{-1}\mathbf{Q}^{-1}\right)[\mathbf{v}]_S,$$

and so $\mathbf{P}^{-1}\mathbf{Q}^{-1}$ acts as the transition matrix from S to B (see Figure 4.8). For example, if $\mathbf{v} = x^2 + 7x + 3$, then

$$[\mathbf{v}]_B = \left(\mathbf{P}^{-1}\mathbf{Q}^{-1}\right)[\mathbf{v}]_S$$

$$= \begin{bmatrix} -6 & -1 & 3 \\ 4 & 2 & -1 \\ 13 & 3 & -6 \end{bmatrix} \begin{bmatrix} 1 & -3 & 2 \\ -1 & 4 & -3 \\ 2 & -5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}.$$

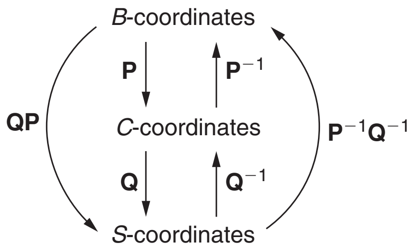

FIGURE 4.8

Transition matrices used to convert among B-, C-, and S-coordinates in $\mathcal{P}_2$

Diagonalization and the Transition Matrix

The matrix $\mathbf{P}$ obtained in the process of diagonalizing an $n \times n$ matrix turns out to be a transition matrix between two different bases for $\mathbb{R}^n$, as we see in our final example.

Example 11

Consider

$$\mathbf{A} = \begin{bmatrix} 14 & -15 & -30 \\ 6 & -7 & -12 \\ 3 & -3 & -7 \end{bmatrix}.$$

A quick calculation produces $p_{\mathbf{A}}(x) = x^3 - 3x - 2 = (x-2)(x+1)^2$. Row reducing $(2\mathbf{I}_3 - \mathbf{A})$ yields a fundamental eigenvector $\mathbf{v}_1 = [5, 2, 1]$. The set $\{\mathbf{v}_1\}$ is a basis for the eigenspace E_2. Similarly, we row reduce $(-1\mathbf{I}_3 - \mathbf{A})$ to obtain fundamental eigenvectors $\mathbf{v}_2 = [1, 1, 0]$ and $\mathbf{v}_3 = [2, 0, 1]$. The set $\{\mathbf{v}_2, \mathbf{v}_3\}$ forms a basis for the eigenspace E_{-1}.

Let $B = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. These vectors are linearly independent (see the remarks before Example 13 in Section 4.4), and thus B is a basis for $\mathbb{R}^3$ by Theorem 4.13. Let S be the standard basis. Then, the transition matrix $\mathbf{P}$ from B to S is given by the matrix whose columns are the

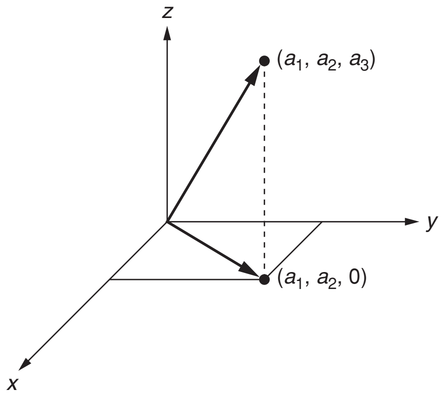

FIGURE 5.3

Projection of $[a_1, a_2, a_3]$ to the xy-plane

consider the mapping $j\colon \mathbb{R}^4 \to \mathbb{R}^4$ given by $j([a_1, a_2, a_3, a_4]) = [0, a_2, 0, a_4]$. This mapping takes each vector in $\mathbb{R}^4$ to a corresponding vector whose first and third coordinates are zero. The functions h and j are both linear operators (see Exercise 5). Such mappings, where at least one of the coordinates is "zeroed out," are examples of **projection mappings**. You can verify that all such mappings are linear operators. (Other types of projection mappings are illustrated in Exercises 6 and 7.)

Example 9

Rotations: Let θ be a fixed angle in $\mathbb{R}^2$, and let $l\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$l\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}.$$

In Exercise 9 you are asked to show that l rotates $[x, y]$ counterclockwise through the angle θ (see Figure 5.4).

Now, let $\mathbf{v}_1 = [x_1, y_1]$ and $\mathbf{v}_2 = [x_2, y_2]$ be two vectors in $\mathbb{R}^2$. Then,

$$\begin{aligned} l(\mathbf{v}_1 + \mathbf{v}_2) &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} (\mathbf{v}_1 + \mathbf{v}_2) \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{v}_1 + \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{v}_2 \\ &= l(\mathbf{v}_1) + l(\mathbf{v}_2). \end{aligned}$$

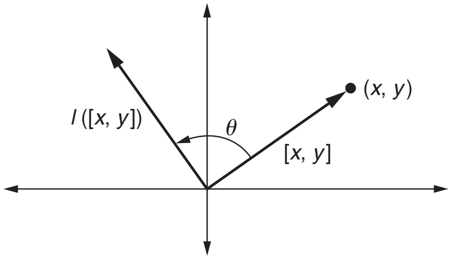

FIGURE 5.4

Counterclockwise rotation of $[x, y]$ through an angle θ in $\mathbb{R}^2$

Similarly, $l(c\mathbf{v}) = cl(\mathbf{v})$, for any $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^2$. Hence, l is a linear operator. ■

Beware! Not all geometric operations are linear operators. Recall that the translation function is not a linear operator!

Multiplication Transformation

The linear operator in Example 9 is actually a special case of the next example, which shows that multiplication by an $m \times n$ matrix is always a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Example 10

Let $\mathbf{A}$ be a given $m \times n$ matrix. We show that the function $f: \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^n$, is a linear transformation. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Then $f(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A}\mathbf{x}_1 + \mathbf{A}\mathbf{x}_2 = f(\mathbf{x}_1) + f(\mathbf{x}_2)$. Also, let $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then, $f(c\mathbf{x}) = \mathbf{A}(c\mathbf{x}) = c(\mathbf{A}\mathbf{x}) = cf(\mathbf{x})$. ■

For a specific example of the multiplication transformation, consider the matrix $\mathbf{A} = \begin{bmatrix} -1 & 4 & 2 \\ 5 & 6 & -3 \end{bmatrix}$. The mapping given by

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -1 & 4 & 2 \\ 5 & 6 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_1 + 4x_2 + 2x_3 \\ 5x_1 + 6x_2 - 3x_3 \end{bmatrix}$$

is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$. In the next section, we will show that the converse of the result in Example 10 also holds; every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is equivalent to multiplication by an appropriate $m \times n$ matrix.

Elementary Properties of Linear Transformations

We now prove some basic properties of linear transformations. From here on, we usually use italicized capital letters, such as "L," to represent linear transformations.

Theorem 5.1 Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Let $\mathbf{0}_{\mathcal{V}}$ be the zero vector in $\mathcal{V}$ and $\mathbf{0}_{\mathcal{W}}$ be the zero vector in $\mathcal{W}$. Then

(1) $L(\mathbf{0}_{\mathcal{V}}) = \mathbf{0}_{\mathcal{W}}$

(2) $L(-\mathbf{v}) = -L(\mathbf{v})$, for all $\mathbf{v} \in \mathcal{V}$

(3) $L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n) = a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2) + \cdots + a_nL(\mathbf{v}_n)$, for all $a_1, \ldots, a_n \in \mathbb{R}$, and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$, for $n \geq 2$.

Proof.

Part (1):

$$\begin{aligned} L(\mathbf{0}_{\mathcal{V}}) &= L(0\mathbf{0}_{\mathcal{V}}) && \text{part (2) of Theorem 4.1, in } \mathcal{V} \\ &= 0L(\mathbf{0}_{\mathcal{V}}) && \text{property (2) of linear transformation} \\ &= \mathbf{0}_{\mathcal{W}} && \text{part (2) of Theorem 4.1, in } \mathcal{W} \end{aligned}$$

Part (2):

$$\begin{aligned} L(-\mathbf{v}) &= L(-1\mathbf{v}) && \text{part (3) of Theorem 4.1, in } \mathcal{V} \\ &= -1(L(\mathbf{v})) && \text{property (2) of linear transformation} \\ &= -L(\mathbf{v}) && \text{part (3) of Theorem 4.1, in } \mathcal{W} \end{aligned}$$

Part (3): (Abridged) This part is proved by induction. We prove the Base Step $(n = 2)$ here and leave the Inductive Step as Exercise 29. For the Base Step, we must show that $L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2)$. But,

$$\begin{aligned} L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) &= L(a_1\mathbf{v}_1) + L(a_2\mathbf{v}_2) && \text{property (1) of linear transformation} \\ &= a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2) && \text{property (2) of linear transformation.} \end{aligned}$$ □

The next theorem asserts that the composition $L_2 \circ L_1$ of linear transformations L_1 and L_2 is again a linear transformation (see Appendix B for a review of composition of functions).

Theorem 5.2 Let $\mathcal{V}_1, \mathcal{V}_2$, and $\mathcal{V}_3$ be vector spaces. Let $L_1: \mathcal{V}_1 \to \mathcal{V}_2$ and $L_2: \mathcal{V}_2 \to \mathcal{V}_3$ be linear transformations. Then $L_2 \circ L_1: \mathcal{V}_1 \to \mathcal{V}_3$ given by $(L_2 \circ L_1)(\mathbf{v}) = L_2(L_1(\mathbf{v}))$, for all $\mathbf{v} \in \mathcal{V}_1$, is a linear transformation.

Proof. (Abridged) To show that $L_2 \circ L_1$ is a linear transformation, we must show that for all $c \in \mathbb{R}$ and $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$,

$$\begin{aligned} (L_2 \circ L_1)(\mathbf{v}_1 + \mathbf{v}_2) &= (L_2 \circ L_1)(\mathbf{v}_1) + (L_2 \circ L_1)(\mathbf{v}_2) \\ \text{and} \quad (L_2 \circ L_1)(c\mathbf{v}) &= c(L_2 \circ L_1)(\mathbf{v}). \end{aligned}$$

The first property holds since

$$\begin{aligned} (L_2 \circ L_1)(\mathbf{v}_1 + \mathbf{v}_2) &= L_2(L_1(\mathbf{v}_1 + \mathbf{v}_2)) \\ &= L_2(L_1(\mathbf{v}_1) + L_1(\mathbf{v}_2)) && \text{because } L_1 \text{ is a linear transformation} \\ &= L_2(L_1(\mathbf{v}_1)) + L_2(L_1(\mathbf{v}_2)) && \text{because } L_2 \text{ is a linear transformation} \\ &= (L_2 \circ L_1)(\mathbf{v}_1) + (L_2 \circ L_1)(\mathbf{v}_2). \end{aligned}$$

We leave the proof of the second property as Exercise 33. □

Example 11

Let L_1 represent the rotation of vectors in $\mathbb{R}^2$ through a fixed angle θ (as in Example 9), and let L_2 represent the reflection of vectors in $\mathbb{R}^2$ through the x-axis. That is, if $\mathbf{v} = [v_1, v_2]$, then

$$L_1(\mathbf{v}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \text{and} \quad L_2(\mathbf{v}) = \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix}.$$

Because L_1 and L_2 are both linear transformations, Theorem 5.2 asserts that

$$L_2(L_1(\mathbf{v})) = L_2\left(\begin{bmatrix} v_1\cos\theta - v_2\sin\theta \\ v_1\sin\theta + v_2\cos\theta \end{bmatrix}\right) = \begin{bmatrix} v_1\cos\theta - v_2\sin\theta \\ -v_1\sin\theta - v_2\cos\theta \end{bmatrix}$$

is also a linear transformation. $L_2 \circ L_1$ represents a rotation of $\mathbf{v}$ through θ followed by a reflection through the x-axis. ■

Theorem 5.2 generalizes naturally to more than two linear transformations. That is, if $L_1, L_2, \ldots, L_k$ are linear transformations and the composition $L_k \circ \cdots \circ L_2 \circ L_1$ makes sense, then $L_k \circ \cdots \circ L_2 \circ L_1$ is also a linear transformation.

Linear Transformations and Subspaces

The final theorem of this section assures us that, under a linear transformation $L: \mathcal{V} \to \mathcal{W}$, subspaces of $\mathcal{V}$ "correspond" to subspaces of $\mathcal{W}$, and vice versa.

Theorem 5.3 Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation.

(1) If $\mathcal{V}'$ is a subspace of $\mathcal{V}$, then $L(\mathcal{V}') = \{L(\mathbf{v}) \mid \mathbf{v} \in \mathcal{V}'\}$, the image of $\mathcal{V}'$ in $\mathcal{W}$, is a subspace of $\mathcal{W}$. In particular, the range of L is a subspace of $\mathcal{W}$.

(2) If $\mathcal{W}'$ is a subspace of $\mathcal{W}$, then $L^{-1}(\mathcal{W}') = \{\mathbf{v} \mid L(\mathbf{v}) \in \mathcal{W}'\}$, the pre-image of $\mathcal{W}'$ in $\mathcal{V}$, is a subspace of $\mathcal{V}$.

We prove part (1) and leave part (2) as Exercise 31.

Proof. Part (1): Suppose that $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and that $\mathcal{V}'$ is a subspace of $\mathcal{V}$. Now, $L(\mathcal{V}')$, the image of $\mathcal{V}'$ in $\mathcal{W}$ (see Figure 5.5), is certainly nonempty (why?). Hence, to show that $L(\mathcal{V}')$ is a subspace of $\mathcal{W}$, we must prove that $L(\mathcal{V}')$ is closed under addition and scalar multiplication.

First, suppose that $\mathbf{w}_1, \mathbf{w}_2 \in L(\mathcal{V}')$. Then, by definition of $L(\mathcal{V}')$, we have $\mathbf{w}_1 = L(\mathbf{v}_1)$ and $\mathbf{w}_2 = L(\mathbf{v}_2)$, for some $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}'$. Then, $\mathbf{w}_1 + \mathbf{w}_2 = L(\mathbf{v}_1) + L(\mathbf{v}_2) = L(\mathbf{v}_1 + \mathbf{v}_2)$ because L is a linear transformation. However, since $\mathcal{V}'$ is a subspace of $\mathcal{V}$, $(\mathbf{v}_1 + \mathbf{v}_2) \in \mathcal{V}'$. Thus, $(\mathbf{w}_1 + \mathbf{w}_2)$ is the image of $(\mathbf{v}_1 + \mathbf{v}_2) \in \mathcal{V}'$, and so $(\mathbf{w}_1 + \mathbf{w}_2) \in L(\mathcal{V}')$. Hence, $L(\mathcal{V}')$ is closed under addition.

Next, suppose that $c \in \mathbb{R}$ and $\mathbf{w} \in L(\mathcal{V}')$. By definition of $L(\mathcal{V}')$, $\mathbf{w} = L(\mathbf{v})$, for some $\mathbf{v} \in \mathcal{V}'$. Then, $c\mathbf{w} = cL(\mathbf{v}) = L(c\mathbf{v})$ since L is a linear transformation. Now, $c\mathbf{v} \in \mathcal{V}'$, because $\mathcal{V}'$ is a subspace of $\mathcal{V}$. Thus, $c\mathbf{w}$ is the image of $c\mathbf{v} \in \mathcal{V}'$, and so $c\mathbf{w} \in L(\mathcal{V}')$. Hence, $L(\mathcal{V}')$ is closed under scalar multiplication. $\square$

Example 12

Let $L: \mathcal{M}_{22} \to \mathbb{R}^3$, where $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = [b, 0, c]$. L is a linear transformation (verify!). By Theorem 5.3, the range of any linear transformation is a subspace of the codomain. Hence, the range of $L = \{[b, 0, c] \mid b, c \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$.

Also, consider the subspace $\mathcal{U}_2 = \left\{\begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \middle| a, b, d \in \mathbb{R}\right\}$ of $\mathcal{M}_{22}$. Then the image of $\mathcal{U}_2$ under L is $\{[b, 0, 0] \mid b \in \mathbb{R}\}$. This image is a subspace of $\mathbb{R}^3$, as Theorem 5.3 asserts. Finally, consider the subspace $\mathcal{W} = \{[b, e, 2b] \mid b, e \in \mathbb{R}\}$ of $\mathbb{R}^3$. The pre-image of $\mathcal{W}$ consists of all

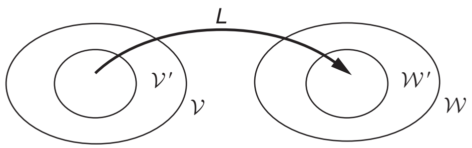

FIGURE 5.5

Subspaces of $\mathcal{V}$ correspond to subspaces of $\mathcal{W}$ under a linear transformation $L: \mathcal{V} \to \mathcal{W}$

matrices in $\mathcal{M}_{22}$ of the form $\begin{bmatrix} a & b \\ 2b & d \end{bmatrix}$. Notice that this pre-image is a subspace of $\mathcal{M}_{22}$, as claimed by Theorem 5.3. ■

New Vocabulary

codomain (of a linear transformation)
composition of linear transformations
contraction (mapping)
dilation (mapping)
domain (of a linear transformation)
identity linear operator
image (of a vector in the domain)
linear operator
linear transformation
pre-image (of a vector in the codomain)
projection (mapping)
range (of a linear transformation)
reflection (mapping)
rotation (mapping)
shear (mapping)
translation (mapping)
zero linear operator

Highlights

- A linear transformation is a function from one vector space to another that preserves the operations of addition and scalar multiplication. That is, under a linear transformation, the image of a linear combination of vectors is the linear combination of the images of the vectors having the same coefficients.
- A linear operator is a linear transformation from a vector space to itself.
- A nontrivial translation of the plane ($\mathbb{R}^2$) or of space ($\mathbb{R}^3$) is never a linear operator, but all of the following are linear operators: contraction (of $\mathbb{R}^n$), dilation (of $\mathbb{R}^n$), reflection of space through the xy-plane (or xz-plane or yz-plane), rotation of the plane about the origin through a given angle θ, projection (of $\mathbb{R}^n$) in which one or more of the coordinates are zeroed out.
- Multiplication of vectors in $\mathbb{R}^n$ on the left by a fixed $m \times n$ matrix $\mathbf{A}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- Multiplying a vector on the left by the matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is equivalent to rotating the vector counterclockwise about the origin through the angle θ.
- Linear transformations always map the zero vector of the domain to the zero vector of the codomain.
- A composition of linear transformations is a linear transformation.
- Under a linear transformation, subspaces of the domain map to subspaces of the codomain, and the pre-image of a subspace of the codomain is a subspace of the domain.

EXERCISES FOR SECTION 5.1

1. Determine which of the following functions are linear transformations. Prove that your answers are correct. Which are linear operators?

   ★(a) $f: \mathbb{R}^2 \to \mathbb{R}^2$ given by $f([x,y]) = [3x - 4y, -x + 2y]$

   ★(b) $h: \mathbb{R}^4 \to \mathbb{R}^4$ given by $h([x_1,x_2,x_3,x_4]) = [x_1 + 2, x_2 - 1, x_3, -3]$

   (c) $k: \mathbb{R}^3 \to \mathbb{R}^3$ given by $k([x_1,x_2,x_3]) = [x_2, x_3, x_1]$

   ★(d) $l: \mathcal{M}_{22} \to \mathcal{M}_{22}$ given by $l\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a - 2c + d & 3b - c \\ -4a & b + c - 3d \end{bmatrix}$

   (e) $n: \mathcal{M}_{22} \to \mathbb{R}$ given by $n\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

   ★(f) $r: \mathcal{P}_3 \to \mathcal{P}_2$ given by $r(ax^3 + bx^2 + cx + d) = (\sqrt[3]{a})x^2 - b^2x + c$

   (g) $s: \mathbb{R}^3 \to \mathbb{R}^3$ given by $s([x_1,x_2,x_3]) = [\cos x_1, \sin x_2, e^{x_3}]$

   ★(h) $t: \mathcal{P}_3 \to \mathbb{R}$ given by $t(a_3x^3 + a_2x^2 + a_1x + a_0) = a_3 + a_2 + a_1 + a_0$

   (i) $u: \mathbb{R}^4 \to \mathbb{R}$ given by $u\left([x_1,x_2,x_3,x_4]\right) = |x_2|$

   ★(j) $v: \mathcal{P}_2 \to \mathbb{R}$ given by $v\left(ax^2 + bx + c\right) = abc$

   ★(k) $g: \mathcal{M}_{32} \to \mathcal{P}_4$ given by $g\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right) = a_{11}x^4 - a_{21}x^2 + a_{31}$

   ★(l) $e: \mathbb{R}^2 \to \mathbb{R}$ given by $e([x,y]) = \sqrt{x^2 + y^2}$

2. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces.

   (a) Show that the identity mapping $i: \mathcal{V} \to \mathcal{V}$ given by $i(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$, is a linear operator.

   (b) Show that the zero mapping $z: \mathcal{V} \to \mathcal{W}$ given by $z(\mathbf{v}) = \mathbf{0}_{\mathcal{W}}$, for all $\mathbf{v} \in \mathcal{V}$, is a linear transformation.

3. Let k be a fixed scalar in $\mathbb{R}$. Show that the mapping $f: \mathbb{R}^n \to \mathbb{R}^n$ given by $f([x_1,x_2,\ldots,x_n]) = k[x_1,x_2,\ldots,x_n]$ is a linear operator.

4. (a) Show that $f: \mathbb{R}^3 \to \mathbb{R}^3$ given by $f([x,y,z]) = [-x,y,z]$ (reflection of a vector through the yz-plane) is a linear operator.

   (b) What mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$ would reflect a vector through the xz-plane? Is it a linear operator? Why or why not?

   (c) What mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ would reflect a vector through the y-axis? through the x-axis? Are these linear operators? Why or why not?

5. Show that the projection mappings $h: \mathbb{R}^3 \to \mathbb{R}^3$ given by $h([a_1,a_2,a_3]) = [a_1,a_2,0]$ and $j: \mathbb{R}^4 \to \mathbb{R}^4$ given by $j([a_1,a_2,a_3,a_4]) = [0,a_2,0,a_4]$ are linear operators.

6. The mapping $f:\mathbb{R}^n \to \mathbb{R}$ given by $f([x_1,x_2,\ldots,x_i,\ldots,x_n]) = x_i$ is another type of projection mapping. Show that f is a linear transformation.

7. Let $\mathbf{x}$ be a fixed nonzero vector in $\mathbb{R}^3$. Show that the mapping $g:\mathbb{R}^3 \to \mathbb{R}^3$ given by $g(\mathbf{y}) = \mathbf{proj}_{\mathbf{x}}\mathbf{y}$ is a linear operator.

8. Let $\mathbf{x}$ be a fixed vector in $\mathbb{R}^n$. Prove that $L:\mathbb{R}^n \to \mathbb{R}$ given by $L(\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ is a linear transformation.

9. Let θ be a fixed angle in the xy-plane. Show that the linear operator $L:\mathbb{R}^2 \to \mathbb{R}^2$ given by $L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$ rotates the vector $[x,y]$ counterclockwise through the angle θ in the plane. (Hint: Consider the vector $[x',y']$, obtained by rotating $[x,y]$ counterclockwise through the angle θ. Let $r = \sqrt{x^2+y^2}$. Then $x = r\cos\alpha$ and $y = r\sin\alpha$, where α is the angle shown in Figure 5.6. Notice that $x' = r(\cos(\theta+\alpha))$ and $y' = r(\sin(\theta+\alpha))$. Then show that $L([x,y]) = [x',y']$.)

10. **(a)** Explain why the mapping $L:\mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

is a linear operator.

(b) Show that the mapping L in part (a) rotates every vector in $\mathbb{R}^3$ about the z-axis through an angle of θ (as measured relative to the xy-plane).

$\star$**(c)** What matrix should be multiplied times $[x,y,z]$ to create the linear operator that rotates $\mathbb{R}^3$ about the y-axis through an angle θ (relative to the xz-plane)? (Hint: When looking down from the positive y-axis toward

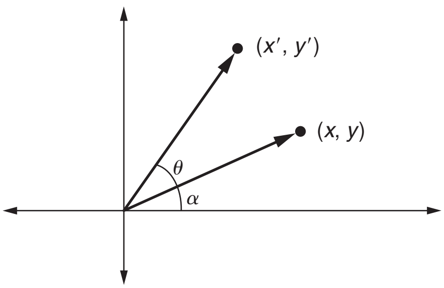

FIGURE 5.6

The vectors $[x,y]$ and $[x',y']$

the xz-plane in a right-handed system, the positive z-axis rotates 90° counterclockwise into the positive x-axis.)

11. **Shears**: Let $f_1, f_2: \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + ky \\ y \end{bmatrix}$$

and

$$f_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ kx + y \end{bmatrix}.$$

The mapping f_1 is called a **shear in the x-direction with factor k**; f_2 is called a **shear in the y-direction with factor k**. The effect of these functions (for $k > 1$) on the vector $[1, 1]$ is shown in Figure 5.7. Show that f_1 and f_2 are linear operators directly, without using Example 10.

12. Let $f: \mathcal{M}_{nn} \to \mathbb{R}$ be given by $f(\mathbf{A}) = \text{trace}(\mathbf{A})$. (The trace is defined in Exercise 14 of Section 1.4.) Prove that f is a linear transformation.

13. Show that the mappings $g, h: \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $g(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T$ and $h(\mathbf{A}) = \mathbf{A} - \mathbf{A}^T$ are linear operators on $\mathcal{M}_{nn}$.

14. **(a)** Show that if $\mathbf{p} \in \mathcal{P}_n$, then the (indefinite integral) function $f: \mathcal{P}_n \to \mathcal{P}_{n+1}$, where $f(\mathbf{p})$ is the vector $\int \mathbf{p}(x)\,dx$ with zero constant term, is a linear transformation.

(b) Show that if $\mathbf{p} \in \mathcal{P}_n$, then the (definite integral) function $g: \mathcal{P}_n \to \mathbb{R}$ given by $g(\mathbf{p}) = \int_a^b \mathbf{p}\,dx$ is a linear transformation, for any fixed $a, b \in \mathbb{R}$.

15. Let $\mathcal{V}$ be the vector space of all functions f from $\mathbb{R}$ to $\mathbb{R}$ that are infinitely differentiable (that is, for which $f^{(n)}$, the nth derivative of f, exists for every

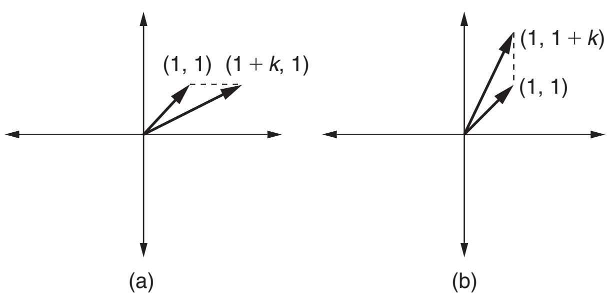

FIGURE 5.7

(a) Shear in the x-direction; (b) shear in the y-direction (both for $k > 0$)

integer $n \geq 1$). Use induction and Theorem 5.2 to show that for any given integer $k \geq 1$, $L: \mathcal{V} \to \mathcal{V}$ given by $L(f) = f^{(k)}$ is a linear operator.

16. Consider the function $f: \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $f(\mathbf{A}) = \mathbf{BA}$, where $\mathbf{B}$ is some fixed $n \times n$ matrix. Show that f is a linear operator.

17. Let $\mathbf{B}$ be a fixed nonsingular matrix in $\mathcal{M}_{nn}$. Show that the mapping $f: \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $f(\mathbf{A}) = \mathbf{B}^{-1}\mathbf{AB}$ is a linear operator.

18. Let a be a fixed real number.
    (a) Let $L: \mathcal{P}_n \to \mathbb{R}$ be given by $L(\mathbf{p}(x)) = \mathbf{p}(a)$. (That is, L evaluates polynomials in $\mathcal{P}_n$ at $x = a$.) Show that L is a linear transformation.
    (b) Let $L: \mathcal{P}_n \to \mathcal{P}_n$ be given by $L(\mathbf{p}(x)) = \mathbf{p}(x + a)$. (For example, when a is positive, L shifts the graph of $\mathbf{p}(x)$ to the *left* by a units.) Prove that L is a linear operator.

19. Let $\mathbf{A}$ be a fixed matrix in $\mathcal{M}_{nn}$. Define $f: \mathcal{P}_n \to \mathcal{M}_{nn}$ by

$$f(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)$$
$$= a_n\mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I}_n.$$

Show that f is a linear transformation.

20. Let $\mathcal{V}$ be the unusual vector space from Example 7 in Section 4.1. Show that $L: \mathcal{V} \to \mathbb{R}$ given by $L(x) = \ln(x)$ is a linear transformation.

21. Let $\mathcal{V}$ be a vector space, and let $\mathbf{x} \neq \mathbf{0}$ be a fixed vector in $\mathcal{V}$. Prove that the translation function $f: \mathcal{V} \to \mathcal{V}$ given by $f(\mathbf{v}) = \mathbf{v} + \mathbf{x}$ is not a linear transformation.

22. Show that if $\mathbf{A}$ is a fixed matrix in $\mathcal{M}_{mn}$ and $\mathbf{y} \neq \mathbf{0}$ is a fixed vector in $\mathbb{R}^m$, then the mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ given by $f(\mathbf{x}) = \mathbf{Ax} + \mathbf{y}$ is not a linear transformation by showing that part (1) of Theorem 5.1 fails for f.

23. Prove that $f: \mathcal{M}_{33} \to \mathbb{R}$ given by $f(\mathbf{A}) = |\mathbf{A}|$ is not a linear transformation. (A similar result is true for $\mathcal{M}_{nn}$, for $n > 1$.)

24. Suppose $L_1: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $L_2: \mathcal{V} \to \mathcal{W}$ is defined by $L_2(\mathbf{v}) = L_1(2\mathbf{v})$. Show that L_2 is a linear transformation.

25. Suppose $L: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear operator and $L([1,0,0]) = [-2,1,0]$, $L([0,1,0]) = [3,-2,1]$, and $L([0,0,1]) = [0,-1,3]$. Find $L([-3,2,4])$. Give a formula for $L([x,y,z])$, for any $[x,y,z] \in \mathbb{R}^3$.

★26. Suppose $L: \mathbb{R}^2 \to \mathbb{R}^2$ is a linear operator and $L(\mathbf{i} + \mathbf{j}) = \mathbf{i} - 3\mathbf{j}$ and $L(-2\mathbf{i} + 3\mathbf{j}) = -4\mathbf{i} + 2\mathbf{j}$. Express $L(\mathbf{i})$ and $L(\mathbf{j})$ as linear combinations of $\mathbf{i}$ and $\mathbf{j}$.

27. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Show that $L(\mathbf{x} - \mathbf{y}) = L(\mathbf{x}) - L(\mathbf{y})$, for all vectors $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

28. Part (3) of Theorem 5.1 assures us that if $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then $L(a\mathbf{v}_1 + b\mathbf{v}_2) = aL(\mathbf{v}_1) + bL(\mathbf{v}_2)$, for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$ and all $a, b \in \mathbb{R}$. Prove that the converse of this statement is true. (Hint: Consider two cases: first $a = b = 1$ and then $b = 0$.)

▶**29.** Finish the proof of part (3) of Theorem 5.1 by doing the Inductive Step.

30. **(a)** Suppose that $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation. Show that if $\{L(\mathbf{v}_1), L(\mathbf{v}_2), \ldots, L(\mathbf{v}_n)\}$ is a linearly independent set of n distinct vectors in $\mathcal{W}$, for some vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set in $\mathcal{V}$.

⋆**(b)** Find a counterexample to the converse of part (a).

▶**31.** Finish the proof of Theorem 5.3 by showing that if $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\mathcal{W}'$ is a subspace of $\mathcal{W}$ with pre-image $L^{-1}(\mathcal{W}')$, then $L^{-1}(\mathcal{W}')$ is a subspace of $\mathcal{V}$.

32. Show that every linear operator $L: \mathbb{R} \to \mathbb{R}$ has the form $L(\mathbf{x}) = c\mathbf{x}$, for some $c \in \mathbb{R}$.

33. Finish the proof of Theorem 5.2 by proving property (2) of a linear transformation for $L_2 \circ L_1$.

34. Let $L_1, L_2: \mathcal{V} \to \mathcal{W}$ be linear transformations. Define $(L_1 \oplus L_2): \mathcal{V} \to \mathcal{W}$ by $(L_1 \oplus L_2)(\mathbf{v}) = L_1(\mathbf{v}) + L_2(\mathbf{v})$ (where the latter addition takes place in $\mathcal{W}$). Also define $(c \odot L_1): \mathcal{V} \to \mathcal{W}$ by $(c \odot L_1)(\mathbf{v}) = c(L_1(\mathbf{v}))$ (where the latter scalar multiplication takes place in $\mathcal{W}$).

(a) Show that $(L_1 \oplus L_2)$ and $(c \odot L_1)$ are linear transformations.

(b) Use the results in part (a) above and part (b) of Exercise 2 to show that the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$ is a vector space under the operations $\oplus$ and $\odot$.

35. Let $L: \mathbb{R}^2 \to \mathbb{R}^2$ be a nonzero linear operator. Show that L maps a line to either a line or a point.

⋆**36.** True or False:

(a) If $L: \mathcal{V} \to \mathcal{W}$ is a function between vector spaces for which $L(c\mathbf{v}) = cL(\mathbf{v})$, then L is a linear transformation.

(b) If $\mathcal{V}$ is an n-dimensional vector space with ordered basis B, then $L: \mathcal{V} \to \mathbb{R}^n$ given by $L(\mathbf{v}) = [\mathbf{v}]_B$ is a linear transformation.

(c) The function $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x, y, z]) = [x+1, y-2, z+3]$ is a linear operator.

(d) If $\mathbf{A}$ is a 4×3 matrix, then $L(\mathbf{v}) = \mathbf{A}\mathbf{v}$ is a linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$.

(e) A linear transformation from $\mathcal{V}$ to $\mathcal{W}$ always maps $\mathbf{0}_{\mathcal{V}}$ to $\mathbf{0}_{\mathcal{W}}$.

(f) If $M_1: \mathcal{V} \to \mathcal{W}$ and $M_2: \mathcal{W} \to \mathcal{X}$ are linear transformations, then $M_1 \circ M_2$ is a well-defined linear transformation.

(g) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the image of any subspace of $\mathcal{V}$ is a subspace of $\mathcal{W}$.

(h) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the pre-image of $\{\mathbf{0}_{\mathcal{W}}\}$ is a subspace of $\mathcal{V}$.

5.2 THE MATRIX OF A LINEAR TRANSFORMATION

In this section, we show that the behavior of any linear transformation $L: \mathcal{V} \to \mathcal{W}$ is determined by its effect on a basis for $\mathcal{V}$. In particular, when $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional and ordered bases for $\mathcal{V}$ and $\mathcal{W}$ are chosen, we can obtain a matrix corresponding to L that is useful in computing images under L. Finally, we investigate how the matrix for L changes as the bases for $\mathcal{V}$ and $\mathcal{W}$ change.

A Linear Transformation Is Determined by Its Action on a Basis

If the action of a linear transformation $L: \mathcal{V} \to \mathcal{W}$ on a basis for $\mathcal{V}$ is known, then the action of L can be computed for all elements of $\mathcal{V}$, as we see in the next example.

Example 1

You can quickly verify that

$$B = \big([0,4,0,1], [-2,5,0,2], [-3,5,1,1], [-1,2,0,1]\big)$$

is an ordered basis for $\mathbb{R}^4$. Now suppose that $L: \mathbb{R}^4 \to \mathbb{R}^3$ is a linear transformation for which

$$L\big([0,4,0,1]\big) = [3,1,2], \qquad L([-2,5,0,2]) = [2,-1,1],$$

$$L([-3,5,1,1]) = [-4,3,0], \quad \text{and} \quad L([-1,2,0,1]) = [6,1,-1].$$

We can use the values of L on B to compute L for other vectors in $\mathbb{R}^4$. For example, let $\mathbf{v} = [-4,14,1,4]$. By using row reduction, we see that $[\mathbf{v}]_B = [2,-1,1,3]$ (verify!). So,

$$\begin{aligned}
L(\mathbf{v}) &= L\big(2[0,4,0,1] - 1[-2,5,0,2] + 1[-3,5,1,1] + 3[-1,2,0,1]\big) \\
&= 2L\big([0,4,0,1]\big) - 1L([-2,5,0,2]) + 1L([-3,5,1,1]) \\
&\quad + 3L([-1,2,0,1]) \\
&= 2[3,1,2] - [2,-1,1] + [-4,3,0] + 3[6,1,-1] \\
&= [18,9,0].
\end{aligned}$$

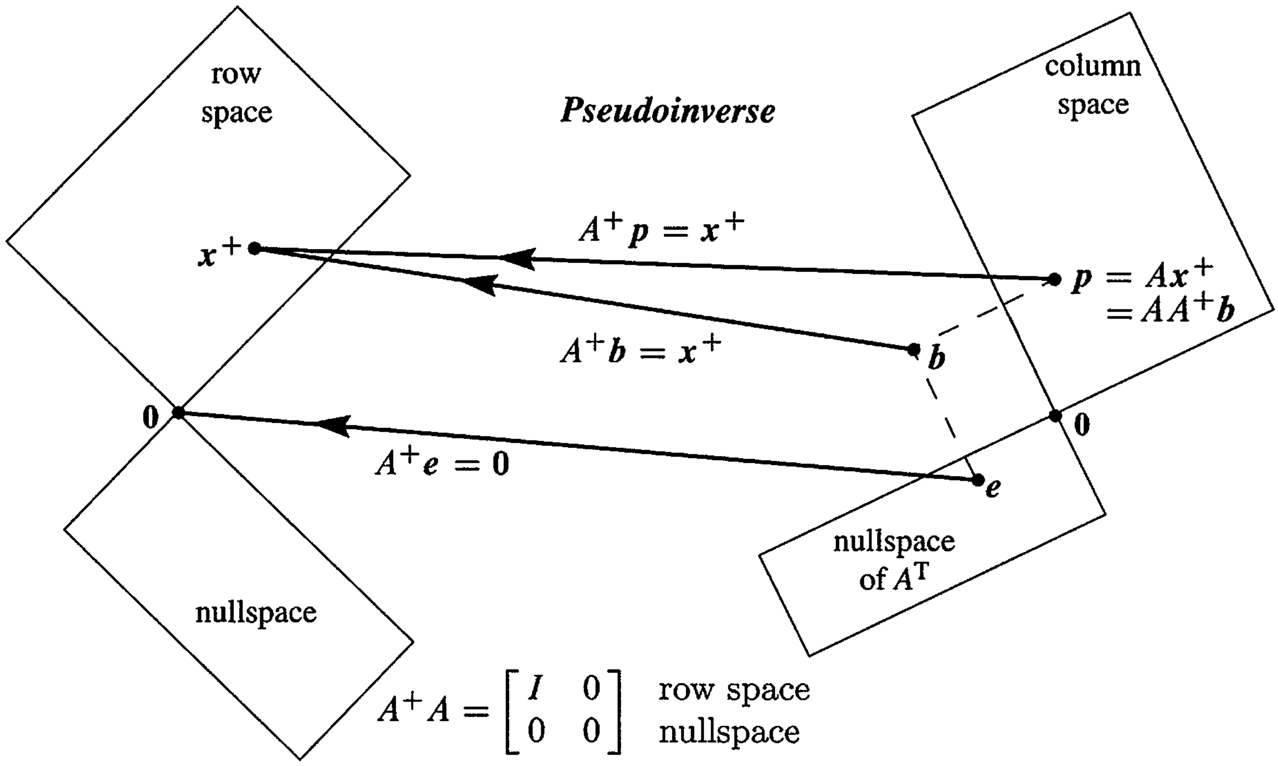

Figure 7.4: Ax^+ in the column space goes back to $A^+Ax^+ = x^+$ in the row space.

Trying for $AA^{-1} = A^{-1}A = I$	AA^+ = projection matrix onto the column space of A A^+A = projection matrix onto the row space of A

Example 4 Find the pseudoinverse of $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$. This matrix is not invertible. The rank is 1. The only singular value is $\sqrt{10}$. That is inverted to $1/\sqrt{10}$ in Σ^+:

$$A^+ = V\Sigma^+U^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{10} & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}.$$

A^+ also has rank 1. Its column space is the row space of A. When A takes $(1,1)$ in the row space to $(4,2)$ in the column space, A^+ does the reverse. $A^+(4,2) = (1,1)$.

Every rank one matrix is a column times a row. With unit vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, that is $A = \sigma \boldsymbol{u}\boldsymbol{v}^T$. Then the best inverse of a rank one matrix is $A^+ = \boldsymbol{v}\boldsymbol{u}^T/\sigma$. The product AA^+ is $\boldsymbol{u}\boldsymbol{u}^T$, the projection onto the line through $\boldsymbol{u}$. The product A^+A is $\boldsymbol{v}\boldsymbol{v}^T$.

Application to least squares Chapter 4 found the best solution $\widehat{\boldsymbol{x}}$ to an unsolvable system $A\boldsymbol{x} = \boldsymbol{b}$. The key equation is $A^TA\widehat{\boldsymbol{x}} = A^T\boldsymbol{b}$, with the assumption that A^TA is invertible. The zero vector was alone in the nullspace.

Now A may have dependent columns (rank $< n$). There can be many solutions to $A^TA\widehat{\boldsymbol{x}} = A^T\boldsymbol{b}$. ***One solution is*** $\boldsymbol{x}^+ = A^+\boldsymbol{b}$ ***from the pseudoinverse.*** We can check that

$A^{\mathrm{T}}AA^+\boldsymbol{b}$, is $A^{\mathrm{T}}\boldsymbol{b}$, because Figure 7.4 shows that $\boldsymbol{e} = \boldsymbol{b} - AA^+\boldsymbol{b}$ is the part of $\boldsymbol{b}$ in the nullspace of A^{T}. Any vector in the nullspace of A could be added to $\boldsymbol{x}^+$, to give another solution $\widehat{x}$ to $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. But $\boldsymbol{x}^+$ will be shorter than any other $\widehat{x}$ (Problem 16):

The shortest least squares solution to $Ax = \boldsymbol{b}$ is $\boldsymbol{x}^+ = A^+\boldsymbol{b}$.

The pseudoinverse A^+ and this best solution $\boldsymbol{x}^+$ are essential in statistics, because experiments often have a matrix A with **dependent columns**.

■ REVIEW OF THE KEY IDEAS ■

1. Diagonalization $S^{-1}AS = \Lambda$ is the same as a change to the eigenvector basis.
2. The SVD chooses an input basis of $\boldsymbol{v}$'s and an output basis of $\boldsymbol{u}$'s. Those orthonormal bases diagonalize A. This is $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$, and in matrix form $A = U\Sigma V^{\mathrm{T}}$.
3. Polar decomposition factors A into QH, rotation UV^{T} times stretching $V\Sigma V^{\mathrm{T}}$.
4. The pseudoinverse $A^+ = V\Sigma^+U^{\mathrm{T}}$ transforms the column space of A back to its row space. A^+A is the identity on the row space (and zero on the nullspace).

■ WORKED EXAMPLES ■

7.3 A If A has rank n (full column rank) then it has a *left inverse* $C = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. This matrix C gives $CA = I$. Explain why the pseudoinverse is $A^+ = C$ in this case. If A has rank m (full row rank) then it has a *right inverse* B with $B = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$. Then $AB = I$. Explain why $A^+ = B$ in this case.

Find B for A_1 and find C for A_2. Find A^+ for all three matrices A_1, A_2, A_3:

$$A_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad A_2 = \begin{bmatrix} 2 & 2 \end{bmatrix} \quad A_3 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}.$$

Solution If A has rank n (independent columns) then $A^{\mathrm{T}}A$ is invertible—this is a key point of Section 4.2. Certainly $C = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ multiplies A to give $CA = I$. In the opposite order, $AC = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is the projection matrix (Section 4.2 again) onto the column space. So C meets the requirements to be A^+: *CA and AC are projections.*

If A has rank m (full row rank) then AA^{T} is invertible. Certainly A multiplies $B = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$ to give $AB = I$. In the opposite order, $BA = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A$ is the projection matrix onto the row space. So B is the pseudoinverse A^+ with rank m.

6.5 Positive Definite Matrices

This section concentrates on ***symmetric matrices that have positive eigenvalues***. If symmetry makes a matrix important, this extra property (*all* $\lambda > 0$) makes it truly special. When we say special, we don't mean rare. Symmetric matrices with positive eigenvalues are at the center of all kinds of applications. They are called ***positive definite***.

The first problem is to recognize these matrices. You may say, just find the eigenvalues and test $\lambda > 0$. That is exactly what we want to avoid. Calculating eigenvalues is work. When the λ's are needed, we can compute them. But if we just want to know that they are positive, there are faster ways. Here are two goals of this section:

- To find *quick tests* on a symmetric matrix that guarantee *positive eigenvalues*.
- To explain important applications of positive definiteness.

The λ's are automatically real because the matrix is symmetric.

Start with 2 by 2. When does $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ ***have*** $\lambda_1 > 0$ ***and*** $\lambda_2 > 0$?

The eigenvalues of A are positive if and only if $a > 0$ ***and*** $ac - b^2 > 0$.

$$A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \text{ is } \textit{not} \text{ positive definite because } ac - b^2 = 1 - 4 < 0$$

$$A_2 = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix} \text{ is positive definite because } a = 1 \text{ and } ac - b^2 = 6 - 4 > 0$$

$$A_3 = \begin{bmatrix} -1 & 2 \\ 2 & -6 \end{bmatrix} \text{ is } \textit{not} \text{ positive definite (even with } \det A = +2) \text{ because } a = -1$$

Notice that we didn't compute the eigenvalues 3 and -1 of A_1. Positive trace $3 - 1 = 2$, negative determinant $(3)(-1) = -3$. And $A_3 = -A_2$ is *negative* definite. The positive eigenvalues for A_2, two negative eigenvalues for A_3.

Proof that the 2 *by* 2 *test is passed when* $\lambda_1 > 0$ *and* $\lambda_2 > 0$. Their product $\lambda_1\lambda_2$ is the determinant so $ac - b^2 > 0$. Their sum is the trace so $a + c > 0$. Then a and c are both positive (if one of them is not positive, $ac - b^2 > 0$ will fail). Problem 1 reverses the reasoning to show that the tests guarantee $\lambda_1 > 0$ and $\lambda_2 > 0$.

This test uses the 1 by 1 determinant a and the 2 by 2 determinant $ac - b^2$. When A is 3 by 3, $\det A > 0$ is the third part of the test. The next test requires *positive pivots*.

The eigenvalues of $A = A^{\mathrm{T}}$ ***are positive if and only if the pivots are positive:***

$$a > 0 \quad \text{and} \quad \frac{ac - b^2}{a} > 0.$$

$a > 0$ is required in both tests. So $ac > b^2$ is also required, for the determinant test and now the pivot. The point is to recognize that ratio as the *second pivot* of A:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \xrightarrow[\text{The multiplier is } b/a]{\text{The first pivot is } a} \begin{bmatrix} a & b \\ 0 & c - \frac{b}{a}b \end{bmatrix} \quad \begin{array}{c}\textbf{\textit{The second pivot is}} \\ c - \dfrac{b^2}{a} = \dfrac{ac - b^2}{a}\end{array}$$

This connects two big parts of linear algebra. ***Positive eigenvalues mean positive pivots and vice versa.*** We gave a proof for symmetric matrices of any size in the last section. The pivots give a quick test for $\lambda > 0$, and they are a lot faster to compute than the eigenvalues. It is very satisfying to see pivots and determinants and eigenvalues come together in this course.

$$A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix} \qquad A_3 = \begin{bmatrix} -1 & 2 \\ 2 & -6 \end{bmatrix}$$

pivots 1 and −3 **(indefinite)**   pivots 1 and 2 **(positive definite)**   pivots −1 and −2 **(negative definite)**

Here is a different way to look at symmetric matrices with positive eigenvalues.

Energy-based Definition

From $Ax = \lambda x$, multiply by x^T to get $x^T Ax = \lambda x^T x$. The right side is a positive λ times a positive number $x^T x = \|x\|^2$. So $x^T Ax$ is positive for any eigenvector.

The new idea is that $x^T Ax$ ***is positive for all nonzero vectors*** x, not just the eigenvectors. In many applications this number $x^T Ax$ (or $\frac{1}{2}x^T Ax$) is the **energy** in the system. The requirement of positive energy gives *another definition* of a positive definite matrix. I think this energy-based definition is the fundamental one.

Eigenvalues and pivots are two equivalent ways to test the new requirement $x^T Ax > 0$.

Definition *A is positive definite if $x^T Ax > 0$ for every nonzero vector x:*

$$x^T Ax = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2 > 0. \tag{1}$$

The four entries a, b, b, c give the four parts of $x^T Ax$. From a and c come the pure squares ax^2 and cy^2. From b and b off the diagonal come the cross terms bxy and byx (the same). Adding those four parts gives $x^T Ax$. This energy-based definition leads to a basic fact:

If A and B are symmetric positive definite, so is $A + B$.

Reason: $x^T(A + B)x$ is simply $x^T Ax + x^T Bx$. Those two terms are positive (for $x \neq \mathbf{0}$) so $A + B$ is also positive definite. The pivots and eigenvalues are not easy to follow when matrices are added, but the energies just add.

$x^{\mathrm{T}}Ax$ also connects with our final way to recognize a positive definite matrix. Start with any matrix R, possibly rectangular. We know that $A = R^{\mathrm{T}}R$ is square and symmetric. More than that, A will be positive definite when R has independent columns:

If the columns of R are independent, then $A = R^{\mathrm{T}}R$ is positive definite.

Again eigenvalues and pivots are not easy. But the number $x^{\mathrm{T}}Ax$ is the same as $x^{\mathrm{T}}R^{\mathrm{T}}Rx$. That is exactly $(Rx)^{\mathrm{T}}(Rx)$—another important proof by parenthesis! That vector Rx is not zero when $x \neq \mathbf{0}$ (this is the meaning of independent columns). Then $x^{\mathrm{T}}Ax$ is the positive number $\|Rx\|^2$ and the matrix A is positive definite.

Let me collect this theory together, into five equivalent statements of positive definiteness. You will see how that key idea connects the whole subject of linear algebra: pivots, determinants, eigenvalues, and least squares (from $R^{\mathrm{T}}R$). Then come the applications.

***When a symmetric matrix has one of these five properties, it has them all*:**

1. All ***n pivots*** are positive.
2. All ***n upper left determinants*** are positive.
3. All ***n eigenvalues*** are positive.
4. $x^{\mathrm{T}}Ax$ is positive except at $x = \mathbf{0}$. This is the ***energy-based*** definition.
5. A equals $R^{\mathrm{T}}R$ for a matrix R with ***independent columns***.

The "upper left determinants" are 1 by 1, 2 by 2, ..., n by n. The last one is the determinant of the complete matrix A. This remarkable theorem ties together the whole linear algebra course—at least for symmetric matrices. We believe that two examples are more helpful than a detailed proof (we nearly have a proof already).

Example 1 Test these matrices A and B for positive definiteness:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 & b \\ -1 & 2 & -1 \\ b & -1 & 2 \end{bmatrix}.$$

Solution The pivots of A are 2 and $\frac{3}{2}$ and $\frac{4}{3}$, all positive. Its upper left determinants are 2 and 3 and 4, all positive. The eigenvalues of A are $2 - \sqrt{2}$ and 2 and $2 + \sqrt{2}$, all positive. That completes tests **1**, **2**, and **3**.

We can write $x^{\mathrm{T}}Ax$ as a sum of three squares. The pivots 2, $\frac{3}{2}$, $\frac{4}{3}$ appear outside the squares. The multipliers $-\frac{1}{2}$ and $-\frac{2}{3}$ from elimination are inside the squares:

$$x^{\mathrm{T}}Ax = 2\left(x_1^2 - x_1x_2 + x_2^2 - x_2x_3 + x_3^2\right) \qquad \textbf{Rewrite with squares}$$

$$= 2\left(x_1 - \tfrac{1}{2}x_2\right)^2 + \tfrac{3}{2}\left(x_2 - \tfrac{2}{3}x_3\right)^2 + \tfrac{4}{3}(x_3)^2. \qquad \textbf{This sum is positive.}$$

I have two candidates to suggest for R. Either one will show that $A = R^{\mathrm{T}}R$ is positive definite. R can be a rectangular first difference matrix, 4 by 3, to produce those second differences $-1, 2, -1$ in A:

$$\boldsymbol{A = R^{\mathrm{T}}R} \qquad \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

The three columns of this R are independent. A is positive definite.

Another R comes from $A = LDL^{\mathrm{T}}$ (the symmetric version of $A = LU$). Elimination gives the pivots $2, \frac{3}{2}, \frac{4}{3}$ in D and the multipliers $-\frac{1}{2}, 0, -\frac{2}{3}$ in L. **Just put $\sqrt{D}$ with L.**

$$LDL^{\mathrm{T}} = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & \\ & 1 & -\frac{2}{3} \\ & & 1 \end{bmatrix} = \begin{array}{l} (L\sqrt{D})(L\sqrt{D})^{\mathrm{T}} = R^{\mathrm{T}}R. \\ \textbf{\textit{R is the Cholesky factor}} \end{array} \qquad (2)$$

This choice of R has square roots (not so beautiful). But it is the only R that is 3 by 3 and upper triangular. It is the "Cholesky factor" of A and it is computed by MATLAB's command $R = \mathsf{chol}(A)$. In applications, the rectangular R is how we build A and this Cholesky R is how we break it apart.

Eigenvalues give the symmetric choice $R = Q\sqrt{\Lambda}Q^{\mathrm{T}}$. This is also successful with $R^{\mathrm{T}}R = Q\Lambda Q^{\mathrm{T}} = A$. All these tests show that the $-1, 2, -1$ matrix A is positive definite.

Now turn to B, where the $(1,3)$ and $(3,1)$ entries move away from 0 to b. This b must not be too large! *The determinant test is easiest.* The 1 by 1 determinant is 2, the 2 by 2 determinant is still 3. The 3 by 3 determinant involves b:

$$\det B = 4 + 2b - 2b^2 = (1+b)(4-2b) \quad \text{must be positive.}$$

At $b = -1$ and $b = 2$ we get $\det B = 0$. *Between $b = -1$ and $b = 2$ the matrix is positive definite.* The corner entry $b = 0$ in the first matrix A was safely between.

Positive Semidefinite Matrices

Often we are at the edge of positive definiteness. The determinant is zero. The smallest eigenvalue is zero. The energy in its eigenvector is $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}0\boldsymbol{x} = 0$. These matrices on the edge are called *positive semidefinite*. Here are two examples (not invertible):

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \text{ are positive semidefinite.}$$

A has eigenvalues 5 and 0. Its upper left determinants are 1 and 0. Its rank is only 1. This matrix A factors into $R^{\mathrm{T}}R$ with **dependent columns** in R:

$$\begin{array}{l} \textbf{Dependent columns} \\ \textbf{Positive semidefinite} \end{array} \quad \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = R^{\mathrm{T}}R.$$

If 4 is increased by any small number, the matrix will become positive definite.

The cyclic B also has zero determinant (computed above when $b = -1$). It is singular. The eigenvector $\boldsymbol{x} = (1,1,1)$ has $B\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{x}^{\mathrm{T}}B\boldsymbol{x} = 0$. Vectors $\boldsymbol{x}$ in all other directions do give positive energy. This B can be written as $R^{\mathrm{T}}R$ in many ways, but R will always have *dependent* columns, with $(1,1,1)$ in its nullspace:

Second differences A
from first differences $R^{\mathrm{T}}R$
Cyclic A **from cyclic** R

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

Positive semidefinite matrices have all $\lambda \geq 0$ and all $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} \geq 0$. Those weak inequalities ($\geq$ instead of $>$) include positive definite matrices and the singular matrices at the edge.

First Application: The Ellipse $ax^2 + 2bxy + cy^2 = 1$

Think of a tilted ellipse $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = 1$. Its center is $(0,0)$, as in Figure 6.7a. Turn it to line up with the coordinate axes (X and Y axes). That is Figure 6.7b. These two pictures show the geometry behind the factorization $A = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$:

1. The tilted ellipse is associated with A. Its equation is $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = 1$.

2. The lined-up ellipse is associated with Λ. Its equation is $\boldsymbol{X}^{\mathrm{T}}\Lambda\boldsymbol{X} = 1$.

3. The rotation matrix that lines up the ellipse is the eigenvector matrix Q.

Example 2 Find the axes of this tilted ellipse $5x^2 + 8xy + 5y^2 = 1$.

Solution Start with the positive definite matrix that matches this equation:

The equation is $\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$. The matrix is $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$.

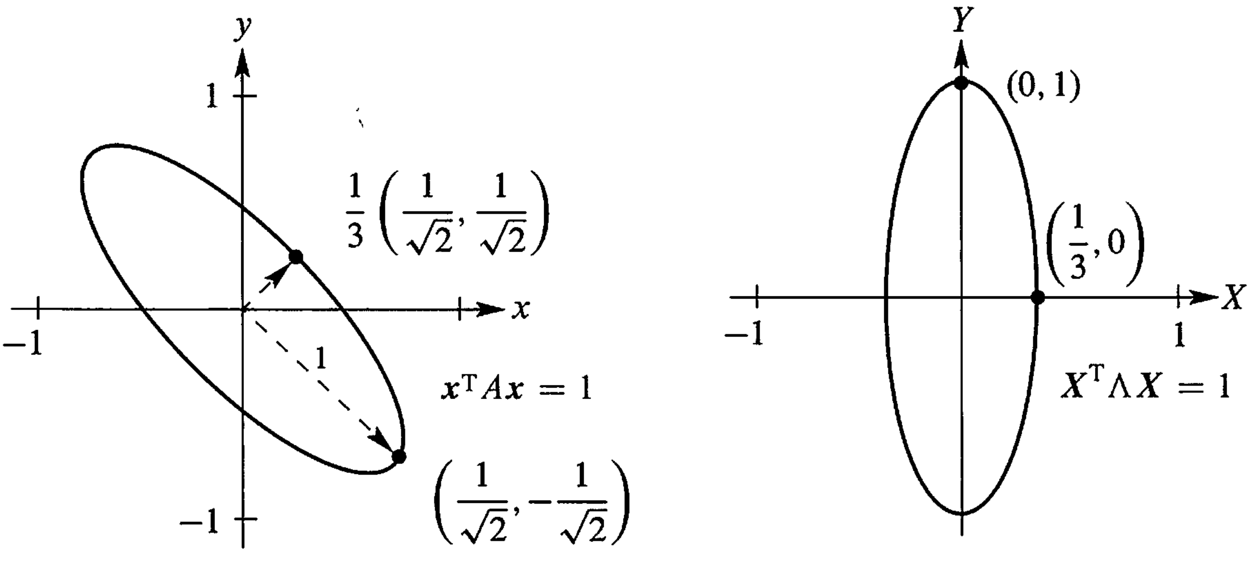

Figure 6.7: The tilted ellipse $5x^2 + 8xy + 5y^2 = 1$. Lined up it is $9X^2 + Y^2 = 1$.

The eigenvectors are $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\-1\end{bmatrix}$. Divide by $\sqrt{2}$ for unit vectors. Then $A = Q\Lambda Q^{\mathrm{T}}$:

$$\begin{array}{l}\textbf{Eigenvectors in } Q\\ \textbf{Eigenvalues 9 and 1}\end{array} \qquad \begin{bmatrix}5 & 4\\ 4 & 5\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\mathbf{9} & 0\\ 0 & \mathbf{1}\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

Now multiply by $\begin{bmatrix}x & y\end{bmatrix}$ on the left and $\begin{bmatrix}x\\y\end{bmatrix}$ on the right to get back to $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$:

$$x^{\mathrm{T}}Ax = \textbf{sum of squares} \quad 5x^2 + 8xy + 5y^2 = 9\left(\frac{x+y}{\sqrt{2}}\right)^2 + 1\left(\frac{x-y}{\sqrt{2}}\right)^2. \qquad (3)$$

The coefficients are not the pivots 5 and 9/5 from D, they are the eigenvalues 9 and 1 from Λ. Inside *these* squares are the eigenvectors $(1,1)/\sqrt{2}$ and $(1,-1)/\sqrt{2}$.

The axes of the tilted ellipse point along the eigenvectors. This explains why $A = Q\Lambda Q^{\mathrm{T}}$ is called the "principal axis theorem"—it displays the axes. Not only the axis directions (from the eigenvectors) but also the axis lengths (from the eigenvalues). To see it all, use capital letters for the new coordinates that line up the ellipse:

$$\textbf{Lined up} \qquad \frac{x+y}{\sqrt{2}} = X \qquad \text{and} \qquad \frac{x-y}{\sqrt{2}} = Y \qquad \text{and} \qquad 9X^2 + Y^2 = 1.$$

The largest value of X^2 is $1/9$. The endpoint of the shorter axis has $X = 1/3$ and $Y = 0$. Notice: The *bigger* eigenvalue λ_1 gives the *shorter* axis, of half-length $1/\sqrt{\lambda_1} = 1/3$. The smaller eigenvalue $\lambda_2 = 1$ gives the greater length $1/\sqrt{\lambda_2} = 1$.

In the xy system, the axes are along the eigenvectors of A. In the XY system, the axes are along the eigenvectors of Λ—the coordinate axes. All comes from $A = Q\Lambda Q^{\mathrm{T}}$.

Suppose $A = Q\Lambda Q^{\mathrm{T}}$ is positive definite, so $\lambda_i > 0$. The graph of $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = 1$ is an ellipse:

$$\begin{bmatrix}x & y\end{bmatrix} Q\Lambda Q^{\mathrm{T}}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}X & Y\end{bmatrix}\Lambda\begin{bmatrix}X\\Y\end{bmatrix} = \lambda_1 X^2 + \lambda_2 Y^2 = 1.$$

The axes point along eigenvectors. The half-lengths are $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$.

$A = I$ gives the circle $x^2 + y^2 = 1$. If one eigenvalue is negative (exchange 4's and 5's in A), we don't have an ellipse. The sum of squares becomes a *difference of squares*: $9X^2 - Y^2 = 1$. This indefinite matrix gives a *hyperbola*. For a negative definite matrix like $A = -I$, with both λ's negative, the graph of $-x^2 - y^2 = 1$ has no points at all.

■ REVIEW OF THE KEY IDEAS ■

1. Positive definite matrices have positive eigenvalues and positive pivots.

2. A quick test is given by the upper left determinants: $a > 0$ and $ac - b^2 > 0$.

3. The graph of $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ is then a "bowl" going up from $\boldsymbol{x}=\mathbf{0}$:

$$\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}=a x^{2}+2 b x y+c y^{2} \text{ is positive except at } (x, y)=(0,0).$$

4. $A=R^{\mathrm{T}} R$ is automatically positive definite if R has independent columns.
5. The ellipse $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}=1$ has its axes along the eigenvectors of A. Lengths $1 / \sqrt{\lambda}$.

■ WORKED EXAMPLES ■

6.5 A The great factorizations of a symmetric matrix are $A=L D L^{\mathrm{T}}$ from pivots and multipliers, and $A=Q \Lambda Q^{\mathrm{T}}$ from eigenvalues and eigenvectors. Show that $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}>0$ for all nonzero $\boldsymbol{x}$ exactly when the pivots and eigenvalues are positive. Try these n by n tests on pascal(6) and ones(6) and hilb(6) and other matrices in MATLAB's gallery.

Solution To prove $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}>0$, put parentheses into $\boldsymbol{x}^{\mathrm{T}} L D L^{\mathrm{T}} \boldsymbol{x}$ and $\boldsymbol{x}^{\mathrm{T}} Q \Lambda Q^{\mathrm{T}} \boldsymbol{x}$:

$$\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}=(L^{\mathrm{T}} \boldsymbol{x})^{\mathrm{T}} D(L^{\mathrm{T}} \boldsymbol{x}) \quad \text{and} \quad \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}=(Q^{\mathrm{T}} \boldsymbol{x})^{\mathrm{T}} \Lambda(Q^{\mathrm{T}} \boldsymbol{x}).$$

If $\boldsymbol{x}$ is nonzero, then $\boldsymbol{y}=L^{\mathrm{T}} \boldsymbol{x}$ and $\boldsymbol{z}=Q^{\mathrm{T}} \boldsymbol{x}$ are nonzero (those matrices are invertible). So $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}=\boldsymbol{y}^{\mathrm{T}} D \boldsymbol{y}=\boldsymbol{z}^{\mathrm{T}} \Lambda \boldsymbol{z}$ becomes a sum of squares and A is shown as positive definite:

$$\begin{array}{llllllll}
\textbf{Pivots} & \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} & = & \boldsymbol{y}^{\mathrm{T}} D \boldsymbol{y} & = & d_{1} y_{1}^{2}+\cdots+d_{n} y_{n}^{2} & > & 0 \\
\textbf{Eigenvalues} & \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} & = & \boldsymbol{z}^{\mathrm{T}} \Lambda \boldsymbol{z} & = & \lambda_{1} z_{1}^{2}+\cdots+\lambda_{n} z_{n}^{2} & > & 0
\end{array}$$

MATLAB has a gallery of unusual matrices (type help gallery) and here are four:

pascal(6) is positive definite because all its pivots are 1 (Worked Example **2.6 A**).

ones(6) is positive *semidefinite* because its eigenvalues are $0,0,0,0,0,6$.

H=hilb(6) is positive definite even though eig(H) shows two eigenvalues very near zero. **Hilbert matrix** $\boldsymbol{x}^{\mathrm{T}} H \boldsymbol{x}=\int_{0}^{1}(x_{1}+x_{2} s+\cdots+x_{6} s^{5})^{2} d s>0$, $H_{i j}=1 /(i+j+1)$.

rand(6)+rand(6)′ can be positive definite or not. Experiments gave only 2 in 20000.

$n=20000$; $p=0$; for $k=1{:}n$, $A=$ rand(6); $p=p+$ all(eig$(A+A')>0$); end, p / n

6.5 B **When is the symmetric block matrix** $M=\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}$ **positive definite?**

Solution Multiply the first row of M by $B^{\mathrm{T}} A^{-1}$ and subtract from the second row, to get a block of zeros. The *Schur complement* $S=C-B^{\mathrm{T}} A^{-1} B$ appears in the corner:

$$\begin{bmatrix} I & 0 \\ -B^{\mathrm{T}} A^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}=\begin{bmatrix} A & B \\ 0 & C-B^{\mathrm{T}} A^{-1} B \end{bmatrix}=\begin{bmatrix} A & B \\ 0 & S \end{bmatrix} \tag{4}$$

Those two blocks A and S must be positive definite. Their pivots are the pivots of M.

6.5 C ***Second application: Test for a minimum.*** Does $F(x, y)$ have a minimum if $\partial F/\partial x = 0$ and $\partial F/\partial y = 0$ at the point $(x, y) = (0, 0)$?

Solution For $f(x)$, the test for a minimum comes from calculus: $df/dx = 0$ and $d^2f/dx^2 > 0$. Moving to two variables x and y produces a symmetric matrix H. It contains the four second derivatives of $F(x, y)$. Positive f'' changes to positive definite H:

Second derivative matrix $$H = \begin{bmatrix} \partial^2 F/\partial x^2 & \partial^2 F/\partial x \partial y \\ \partial^2 F/\partial y \partial x & \partial^2 F/\partial y^2 \end{bmatrix}$$

$F(x, y)$ has a minimum if H is positive definite. Reason: H reveals the important terms $ax^2 + 2bxy + cy^2$ near $(x, y) = (0, 0)$. The second derivatives of F are $2a, 2b, 2b, 2c$!

6.5 D Find the eigenvalues of the $-1, 2, -1$ tridiagonal n by n matrix K (my favorite).

Solution The best way is to guess λ and $\boldsymbol{x}$. Then check $K\boldsymbol{x} = \lambda\boldsymbol{x}$. Guessing could not work for most matrices, but special cases are a big part of mathematics (pure and applied).

The key is hidden in a differential equation. The second difference matrix K is like a *second derivative*, and those eigenvalues are much easier to see:

Eigenvalues $\lambda_1, \lambda_2, \ldots$
Eigenfunctions $y_1, y_2, \ldots$
$$\frac{d^2y}{dx^2} = \lambda y(x) \quad \textbf{with} \quad \begin{matrix} y(0) = 0 \\ y(1) = 0 \end{matrix} \tag{5}$$

Try $y = \sin cx$. Its second derivative is $y'' = -c^2 \sin cx$. So the eigenvalue will be $\lambda = -c^2$, provided $y(x)$ satisfies the end point conditions $y(0) = 0 = y(1)$.

Certainly $\sin 0 = 0$ (this is where cosines are eliminated by $\cos 0 = 1$). At $x = 1$, we need $y(1) = \sin c = 0$. The number c must be $k\pi$, a multiple of π, and λ is $-c^2$:

Eigenvalues $\lambda = -k^2\pi^2$
Eigenfunctions $y = \sin k\pi x$
$$\frac{d^2}{dx^2} \sin k\pi x = -k^2\pi^2 \sin k\pi x. \tag{6}$$

Now we go back to the matrix K and guess its eigenvectors. They come from $\sin k\pi x$ at n points $x = h, 2h, \ldots, nh$, equally spaced between 0 and 1. The spacing Δx is $h = 1/(n+1)$, so the $(n+1)$st point comes out at $(n+1)h = 1$. Multiply that sine vector s by K:

Eigenvector of K = sine vector s
$$\begin{aligned} Ks &= \lambda s = (2 - 2\cos k\pi h)\, s \\ s &= (\sin k\pi h, \ldots, \sin nk\pi h). \end{aligned} \tag{7}$$

I will leave that multiplication $Ks = \lambda s$ as a challenge problem. Notice what is important:

1. All eigenvalues $2 - 2\cos k\pi h$ are positive and K is positive definite.

2. The **sine matrix** S has orthogonal columns = eigenvectors $s_1, \ldots, s_n$ of K.

Discrete Sine Transform
The j, k entry is $\sin jk\pi h$

$$S = \begin{bmatrix} \sin \pi h & & \sin k\pi h & \\ \vdots & \cdots & \vdots & \cdots \\ \sin n\pi h & & \sin nk\pi h & \end{bmatrix}$$

Those eigenvectors are orthogonal just like the eigenfunctions: $\int_0^1 \sin j\pi x \sin k\pi x \, dx = 0$.

Problem Set 6.5

Problems 1–13 are about tests for positive definiteness.

1 Suppose the 2 by 2 tests $a > 0$ and $ac - b^2 > 0$ are passed. Then $c > b^2/a$ is also positive.

(i) λ_1 and λ_2 have the *same sign* because their product $\lambda_1\lambda_2$ equals ____ .

(i) That sign is positive because $\lambda_1 + \lambda_2$ equals ____ .

Conclusion: The tests $a > 0, ac - b^2 > 0$ guarantee positive eigenvalues λ_1, λ_2.

2 Which of A_1, A_2, A_3, A_4 has two positive eigenvalues? Use the test, don't compute the λ's. Find an x so that $x^{\mathrm{T}} A_1 x < 0$, so A_1 fails the test.

$$A_1 = \begin{bmatrix} 5 & 6 \\ 6 & 7 \end{bmatrix} \quad A_2 = \begin{bmatrix} -1 & -2 \\ -2 & -5 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 & 10 \\ 10 & 100 \end{bmatrix} \quad A_4 = \begin{bmatrix} 1 & 10 \\ 10 & 101 \end{bmatrix}.$$

3 For which numbers b and c are these matrices positive definite?

$$A = \begin{bmatrix} 1 & b \\ b & 9 \end{bmatrix} \quad A = \begin{bmatrix} 2 & 4 \\ 4 & c \end{bmatrix} \quad A = \begin{bmatrix} c & b \\ b & c \end{bmatrix}.$$

With the pivots in D and multiplier in L, factor each A into LDL^{T}.

4 What is the quadratic $f = ax^2 + 2bxy + cy^2$ for each of these matrices? Complete the square to write f as a sum of one or two squares $d_1(\quad)^2 + d_2(\quad)^2$.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}.$$

5 Write $f(x, y) = x^2 + 4xy + 3y^2$ as a *difference* of squares and find a point (x, y) where f is negative. The minimum is not at $(0, 0)$ even though f has positive coefficients.

6 The function $f(x, y) = 2xy$ certainly has a saddle point and not a minimum at $(0, 0)$. What symmetric matrix A produces this f? What are its eigenvalues?

7 Test to see if $R^{\mathrm{T}}R$ is positive definite in each case:

$$R = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

8 The function $f(x, y) = 3(x + 2y)^2 + 4y^2$ is positive except at $(0, 0)$. What is the matrix in $f = [x \;\; y]A[x \;\; y]^{\mathrm{T}}$? Check that the pivots of A are 3 and 4.

9 Find the 3 by 3 matrix A and its pivots, rank, eigenvalues, and determinant:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 4(x_1 - x_2 + 2x_3)^2.$$

10 Which 3 by 3 symmetric matrices A and B produce these quadratics?

$$x^{\mathrm{T}}Ax = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3). \quad \text{Why is } A \text{ positive definite?}$$
$$x^{\mathrm{T}}Bx = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_1x_3 - x_2x_3). \quad \text{Why is } B \text{ semidefinite?}$$

11 Compute the three upper left determinants of A to establish positive definiteness. Verify that their ratios give the second and third pivots.

$$\textbf{Pivots = ratios of determinants} \qquad A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 3 & 8 \end{bmatrix}.$$

12 For what numbers c and d are A and B positive definite? Test the 3 determinants:

$$A = \begin{bmatrix} c & 1 & 1 \\ 1 & c & 1 \\ 1 & 1 & c \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & d & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

13 Find a matrix with $a > 0$ and $c > 0$ and $a + c > 2b$ that has a negative eigenvalue.

Problems 14–20 are about applications of the tests.

14 *If A is positive definite then A^{-1} is positive definite.* Best proof: The eigenvalues of A^{-1} are positive because ______. *Second proof* (only for 2 by 2):

$$\text{The entries of } A^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix} \quad \text{pass the determinant tests} \quad \underline{\qquad}.$$

15 *If A and B are positive definite, their sum $A + B$ is positive definite.* Pivots and eigenvalues are not convenient for $A + B$. Better to prove $x^{\mathrm{T}}(A + B)x > 0$. Or if $A = R^{\mathrm{T}}R$ and $B = S^{\mathrm{T}}S$, show that $A + B = [\,R \;\; S\,]^{\mathrm{T}} \begin{bmatrix} R \\ S \end{bmatrix}$ with independent columns.

16 A positive definite matrix cannot have a zero (or even worse, a negative number) on its diagonal. Show that this matrix fails to have $x^{\mathrm{T}}Ax > 0$:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 4 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ is not positive when } (x_1, x_2, x_3) = (\quad, \quad, \quad).$$

17 A diagonal entry a_{jj} of a symmetric matrix cannot be smaller than all the λ's. If it were, then $A - a_{jj}I$ would have _____ eigenvalues and would be positive definite. But $A - a_{jj}I$ has a _____ on the main diagonal.

18 If $Ax = \lambda x$ then $x^{\mathrm{T}}Ax =$ _____. If $x^{\mathrm{T}}Ax > 0$, prove that $\lambda > 0$.

19 Reverse Problem 18 to show that *if all* $\lambda > 0$ *then* $x^{\mathrm{T}}Ax > 0$. We must do this for *every* nonzero x, not just the eigenvectors. So write x as a combination of the eigenvectors and *explain why all "cross terms" are* $x_i^{\mathrm{T}}x_j = 0$. Then $x^{\mathrm{T}}Ax$ is

$$(c_1x_1 + \cdots + c_nx_n)^{\mathrm{T}}(c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n) = c_1^2\lambda_1x_1^{\mathrm{T}}x_1 + \cdots + c_n^2\lambda_nx_n^{\mathrm{T}}x_n > 0.$$

20 Give a quick reason why each of these statements is true:

(a) Every positive definite matrix is invertible.

(b) The only positive definite projection matrix is $P = I$.

(c) A diagonal matrix with positive diagonal entries is positive definite.

(d) A symmetric matrix with a positive determinant might not be positive definite!

Problems 21–24 use the eigenvalues; Problems 25–27 are based on pivots.

21 For which s and t do A and B have all $\lambda > 0$ (therefore positive definite)?

$$A = \begin{bmatrix} s & -4 & -4 \\ -4 & s & -4 \\ -4 & -4 & s \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} t & 3 & 0 \\ 3 & t & 4 \\ 0 & 4 & t \end{bmatrix}.$$

22 From $A = Q\Lambda Q^{\mathrm{T}}$ compute the positive definite symmetric square root $Q\Lambda^{1/2}Q^{\mathrm{T}}$ of each matrix. Check that this square root gives $R^2 = A$:

$$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}.$$

23 You may have seen the equation for an ellipse as $x^2/a^2 + y^2/b^2 = 1$. What are a and b when the equation is written $\lambda_1x^2 + \lambda_2y^2 = 1$? The ellipse $9x^2 + 4y^2 = 1$ has axes with half-lengths $a =$ _____ and $b =$ _____.

24 Draw the tilted ellipse $x^2 + xy + y^2 = 1$ and find the half-lengths of its axes from the eigenvalues of the corresponding matrix A.

25 With positive pivots in D, the factorization $A = LDL^{\mathrm{T}}$ becomes $L\sqrt{D}\sqrt{D}L^{\mathrm{T}}$. (Square roots of the pivots give $D = \sqrt{D}\sqrt{D}$.) Then $C = \sqrt{D}L^{\mathrm{T}}$ yields the ***Cholesky factorization*** $A = C^{\mathrm{T}}C$ which is "symmetrized $L\,U$":

$$\text{From} \quad C = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{find } A. \qquad \text{From} \quad A = \begin{bmatrix} 4 & 8 \\ 8 & 25 \end{bmatrix} \quad \text{find } C = \mathbf{chol}(A).$$

26 In the Cholesky factorization $A = C^{\mathrm{T}}C$, with $C^{\mathrm{T}} = L\sqrt{D}$, the square roots of the pivots are on the diagonal of C. Find C (upper triangular) for

$$A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 8 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 7 \end{bmatrix}.$$

27 The symmetric factorization $A = LDL^{\mathrm{T}}$ means that $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}LDL^{\mathrm{T}}\boldsymbol{x}$:

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1 & 0 \\ b/a & 1 \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & (ac-b^2)/a \end{bmatrix}\begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

The left side is $ax^2 + 2bxy + cy^2$. The right side is $a\left(x + \frac{b}{a}y\right)^2 +$ ____ y^2. The second pivot completes the square! Test with $a = 2, b = 4, c = 10$.

28 Without multiplying $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, find

(a) the determinant of A (b) the eigenvalues of A
(c) the eigenvectors of A (d) a reason why A is symmetric positive definite.

29 For $F_1(x, y) = \frac{1}{4}x^4 + x^2y + y^2$ and $F_2(x, y) = x^3 + xy - x$ find the second derivative matrices H_1 and H_2:

$$\textbf{Test for minimum}\quad H = \begin{bmatrix} \partial^2F/\partial x^2 & \partial^2F/\partial x\partial y \\ \partial^2F/\partial y\partial x & \partial^2F/\partial y^2 \end{bmatrix} \text{ is positive definite}$$

H_1 is positive definite so F_1 is concave up ($=$ convex). Find the minimum point of F_1 and the saddle point of F_2 (look only where first derivatives are zero).

30 The graph of $z = x^2 + y^2$ is a bowl opening upward. *The graph* of $z = x^2 - y^2$ *is a saddle*. The graph of $z = -x^2 - y^2$ is a bowl opening downward. What is a test on a, b, c for $z = ax^2 + 2bxy + cy^2$ to have a saddle point at $(0, 0)$?

31 Which values of c give a bowl and which c give a saddle point for the graph of $z = 4x^2 + 12xy + cy^2$? Describe this graph at the borderline value of c.

Challenge Problems

32 A *group* of nonsingular matrices includes AB and A^{-1} if it includes A and B. "Products and inverses stay in the group." Which of these are groups (as in 2.7.37)?

Invent a "subgroup" of two of these groups (not I by itself $=$ the smallest group).

(a) Positive definite symmetric matrices A.

(b) Orthogonal matrices Q.

(c) All exponentials e^{tA} of a fixed matrix A.

(d) Matrices P with positive eigenvalues.

(e) Matrices D with determinant 1.

33 When A and B are symmetric positive definite, AB might not even be symmetric. But its eigenvalues are still positive. Start from $AB\boldsymbol{x} = \lambda\boldsymbol{x}$ and take dot products with $B\boldsymbol{x}$. Then prove $\lambda > 0$.

34 Write down the 5 by 5 sine matrix S from Worked Example **6.5 D**, containing the eigenvectors of K when $n = 5$ and $h = 1/6$. Multiply K times S to see the five positive eigenvalues.

Their sum should equal the trace 10. Their product should be $\det K = 6$.

35 Suppose C is positive definite (so $\boldsymbol{y}^{\mathrm{T}}C\boldsymbol{y} > 0$ whenever $\boldsymbol{y} \neq \boldsymbol{0}$) and A has independent columns (so $A\boldsymbol{x} \neq \boldsymbol{0}$ whenever $\boldsymbol{x} \neq \boldsymbol{0}$). Apply the energy test to $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}CA\boldsymbol{x}$ to show that $A^{\mathrm{T}}CA$ is positive definite: *the crucial matrix in engineering.*

8.3 Markov Matrices, Population, and Economics

This section is about ***positive matrices***: every $a_{ij} > 0$. The key fact is quick to state: ***The largest eigenvalue is real and positive and so is its eigenvector.*** In economics and ecology and population dynamics and random walks, that fact leads a long way:

$$\textbf{Markov} \quad \lambda_{\max} = 1 \qquad \textbf{Population} \quad \lambda_{\max} > 1 \qquad \textbf{Consumption} \quad \lambda_{\max} < 1$$

$\lambda_{\max}$ controls the powers of A. We will see this first for $\lambda_{\max} = 1$.

Markov Matrices

Suppose we multiply a positive vector $\boldsymbol{u}_0 = (a, 1-a)$ again and again by this A:

Markov matrix $$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \qquad \boldsymbol{u}_1 = A\boldsymbol{u}_0 \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$$

After k steps we have $A^k\boldsymbol{u}_0$. The vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \ldots$ will approach a "*steady state*" $\boldsymbol{u}_\infty = (.6, .4)$. This final outcome does not depend on the starting vector: ***For every $\boldsymbol{u}_0$ we converge to the same $\boldsymbol{u}_\infty$.*** The question is why.

The steady state equation $A\boldsymbol{u}_\infty = \boldsymbol{u}_\infty$ makes $\boldsymbol{u}_\infty$ ***an eigenvector with eigenvalue* 1**:

Steady state $$\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} .6 \\ .4 \end{bmatrix} = \begin{bmatrix} .6 \\ .4 \end{bmatrix}.$$

Multiplying by A does not change $\boldsymbol{u}_\infty$. But this does not explain why all vectors $\boldsymbol{u}_0$ lead to $\boldsymbol{u}_\infty$. Other examples might have a steady state, but it is not necessarily attractive:

Not Markov $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ has the unattractive steady state $B\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

In this case, the starting vector $\boldsymbol{u}_0 = (0, 1)$ will give $\boldsymbol{u}_1 = (0, 2)$ and $\boldsymbol{u}_2 = (0, 4)$. The second components are doubled. In the language of eigenvalues, B has $\lambda = 1$ but also $\lambda = 2$— this produces instability. The component of $\boldsymbol{u}$ along that unstable eigenvector is multiplied by λ, and $|\lambda| > 1$ means blowup.

This section is about two special properties of A that guarantee a stable steady state. These properties define a ***Markov matrix***, and A above is one particular example:

Markov matrix	**1. *Every entry of A is nonnegative.*** **2. *Every column of A adds to* 1.**

B did not have Property **2**. When A is a Markov matrix, two facts are immediate:

1. Multiplying a nonnegative $\boldsymbol{u}_0$ by A produces a nonnegative $\boldsymbol{u}_1 = A\boldsymbol{u}_0$.

2. If the components of $\boldsymbol{u}_0$ add to 1, so do the components of $\boldsymbol{u}_1 = A\boldsymbol{u}_0$.

Reason: The components of $\boldsymbol{u}_0$ add to 1 when $\begin{bmatrix}1 & \cdots & 1\end{bmatrix}\boldsymbol{u}_0 = 1$. This is true for each column of A by Property 2. Then by matrix multiplication $[1 \ \ldots \ 1]A = [1 \ \ldots \ 1]$:

$$\textbf{Components of } A\boldsymbol{u}_0 \textbf{ add to 1} \qquad [1 \ \cdots \ 1]A\boldsymbol{u}_0 = [1 \ \cdots \ 1]\boldsymbol{u}_0 = 1.$$

The same facts apply to $\boldsymbol{u}_2 = A\boldsymbol{u}_1$ and $\boldsymbol{u}_3 = A\boldsymbol{u}_2$. ***Every vector $A^k\boldsymbol{u}_0$ is nonnegative with components adding to*** 1. These are "***probability vectors.***" The limit $\boldsymbol{u}_\infty$ is also a probability vector—but we have to prove that there is a limit. We will show that $\lambda_{\max} = 1$ for a positive Markov matrix.

Example 1 The fraction of rental cars in Denver starts at $\frac{1}{50} = .02$. The fraction outside Denver is .98. Every month, 80% of the Denver cars stay in Denver (and 20% leave). Also 5% of the outside cars come in (95% stay outside). This means that the fractions $\boldsymbol{u}_0 = (.02, .98)$ are multiplied by A:

$$\textbf{First month} \qquad A = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix} \quad \text{leads to} \quad \boldsymbol{u}_1 = A\boldsymbol{u}_0 = A\begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .065 \\ .935 \end{bmatrix}.$$

Notice that $.065 + .935 = 1$. All cars are accounted for. Each step multiplies by A:

$$\textbf{Next month} \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = (.09875,\ .90125). \text{ This is } A^2\boldsymbol{u}_0.$$

All these vectors are positive because A is positive. Each vector $\boldsymbol{u}_k$ will have its components adding to 1. The first component has grown from .02 and cars are moving toward Denver. What happens in the long run?

This section involves powers of matrices. The understanding of A^k was our first and best application of diagonalization. Where A^k can be complicated, the diagonal matrix Λ^k is simple. The eigenvector matrix S connects them: A^k equals $S\Lambda^k S^{-1}$. The new application to Markov matrices uses the eigenvalues (in Λ) and the eigenvectors (in S). We will show that $\boldsymbol{u}_\infty$ is an eigenvector corresponding to $\lambda = 1$.

Since every column of A adds to 1, nothing is lost or gained. We are moving rental cars or populations, and no cars or people suddenly appear (or disappear). The fractions add to 1 and the matrix A keeps them that way. The question is how they are distributed after k time periods—which leads us to A^k.

Solution $A^k\boldsymbol{u}_0$ gives the fractions in and out of Denver after k steps. We diagonalize A to understand A^k. The eigenvalues are $\lambda = 1$ and .75 (the trace is 1.75).

$$Ax = \lambda x \qquad A\begin{bmatrix} .2 \\ .8 \end{bmatrix} = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix} \qquad \textbf{and} \qquad A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = .75\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The starting vector $\boldsymbol{u}_0$ combines $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, in this case with coefficients 1 and .18:

$$\textbf{Combination of eigenvectors} \qquad \boldsymbol{u}_0 = \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Now multiply by A to find $\boldsymbol{u}_1$. The eigenvectors are multiplied by $\lambda_1 = 1$ and $\lambda_2 = .75$:

$$\textbf{Each } \boldsymbol{x} \textbf{ is multiplied by } \boldsymbol{\lambda} \qquad \boldsymbol{u}_1 = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)(.18)\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Every month, another .75 multiplies the vector x_2. The eigenvector x_1 is unchanged:

After k steps $$u_k = A^k u_0 = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)^k(.18)\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

This equation reveals what happens. ***The eigenvector x_1 with $\lambda = 1$ is the steady state.*** The other eigenvector x_2 disappears because $|\lambda| < 1$. The more steps we take, the closer we come to $u_\infty = (.2, .8)$. In the limit, $\frac{2}{10}$ of the cars are in Denver and $\frac{8}{10}$ are outside. This is the pattern for Markov chains, even starting from $u_0 = (0, 1)$:

If A is a *positive* Markov matrix (entries $a_{ij} > 0$, each column adds to **1**), then $\lambda_1 = 1$ is larger than any other eigenvalue. The eigenvector x_1 is the ***steady state***:

$$u_k = x_1 + c_2(\lambda_2)^k x_2 + \cdots + c_n(\lambda_n)^k x_n \quad \textit{always approaches} \quad u_\infty = x_1.$$

The first point is to see that $\lambda = 1$ is an eigenvalue of A. *Reason*: Every column of $A - I$ adds to $1 - 1 = 0$. The rows of $A - I$ add up to the zero row. Those rows are linearly dependent, so $A - I$ is singular. Its determinant is zero and $\lambda = 1$ is an eigenvalue.

The second point is that no eigenvalue can have $|\lambda| > 1$. With such an eigenvalue, the powers A^k would grow. But A^k is also a Markov matrix! A^k has nonnegative entries still adding to 1—and that leaves no room to get large.

A lot of attention is paid to the possibility that another eigenvalue has $|\lambda| = 1$.

Example 2 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has no steady state because $\lambda_2 = -1$.

This matrix sends all cars from inside Denver to outside, and vice versa. The powers A^k alternate between A and I. The second eigenvector $x_2 = (-1, 1)$ will be multiplied by $\lambda_2 = -1$ at every step—and does not become smaller: No steady state.

Suppose the entries of A or any power of A are all *positive*—zero is not allowed. In this "regular" or "primitive" case, $\lambda = 1$ is strictly larger than any other eigenvalue. The powers A^k approach the rank one matrix that has the steady state in every column.

Example 3 (**"Everybody moves"**) Start with three groups. At each time step, half of group 1 goes to group 2 and the other half goes to group 3. The other groups also *split in half and move*. Take one step from the starting populations p_1, p_2, p_3:

New populations $$u_1 = Au_0 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}p_2 + \frac{1}{2}p_3 \\ \frac{1}{2}p_1 + \frac{1}{2}p_3 \\ \frac{1}{2}p_1 + \frac{1}{2}p_2 \end{bmatrix}.$$

A is a Markov matrix. Nobody is born or lost. A contains zeros, which gave trouble in Example 2. But after two steps in this new example, the zeros disappear from A^2:

Two-step matrix $$u_2 = A^2 u_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}.$$

The eigenvalues of A are $\lambda_1 = 1$ (because A is Markov) and $\lambda_2 = \lambda_3 = -\frac{1}{2}$. For $\lambda = 1$, ***the eigenvector*** $\boldsymbol{x}_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ***will be the steady state.*** When three equal populations split in half and move, the populations are again equal. Starting from $\boldsymbol{u}_0 = (8, 16, 32)$, the Markov chain approaches its steady state:

$$\boldsymbol{u}_0 = \begin{bmatrix} 8 \\ 16 \\ 32 \end{bmatrix} \qquad \boldsymbol{u}_1 = \begin{bmatrix} 24 \\ 20 \\ 12 \end{bmatrix} \qquad \boldsymbol{u}_2 = \begin{bmatrix} 16 \\ 18 \\ 22 \end{bmatrix} \qquad \boldsymbol{u}_3 = \begin{bmatrix} 20 \\ 19 \\ 17 \end{bmatrix}.$$

The step to $\boldsymbol{u}_4$ will split some people in half. This cannot be helped. The total population is $8 + 16 + 32 = 56$ at every step. The steady state is 56 times $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. You can see the three populations approaching, but never reaching, their final limits $56/3$.

Challenge Problem 6.7.16 created a Markov matrix A from the number of links between websites. The steady state $\boldsymbol{u}$ will give the Google rankings. ***Google finds*** $\boldsymbol{u}_\infty$ ***by a random walk that follows links*** (*random surfing*). That eigenvector comes from counting the fraction of visits to each website—a quick way to compute the steady state.

The size $|\lambda_2|$ of the next largest eigenvalue controls the speed of convergence to steady state.

Perron-Frobenius Theorem

One matrix theorem dominates this subject. The Perron-Frobenius Theorem applies when all $a_{ij} \geq 0$. There is no requirement that columns add to 1. We prove the neatest form, when all $a_{ij} > 0$.

Perron-Frobenius for $A > 0$ ***All numbers in $A\boldsymbol{x} = \lambda_{\max}\boldsymbol{x}$ are strictly positive.***

Proof The key idea is to look at all numbers t such that $A\boldsymbol{x} \geq t\boldsymbol{x}$ for some nonnegative vector $\boldsymbol{x}$ (other than $\boldsymbol{x} = \boldsymbol{0}$). We are allowing inequality in $A\boldsymbol{x} \geq t\boldsymbol{x}$ in order to have many positive candidates t. For the largest value $t_{\max}$ (which is attained), we will show that ***equality holds***: $A\boldsymbol{x} = t_{\max}\boldsymbol{x}$.

Otherwise, if $A\boldsymbol{x} \geq t_{\max}\boldsymbol{x}$ is not an equality, multiply by A. Because A is positive that produces a strict inequality $A^2\boldsymbol{x} > t_{\max}A\boldsymbol{x}$. Therefore the positive vector $\boldsymbol{y} = A\boldsymbol{x}$ satisfies $A\boldsymbol{y} > t_{\max}\boldsymbol{y}$, and $t_{\max}$ could be increased. This contradiction forces the equality $A\boldsymbol{x} = t_{\max}\boldsymbol{x}$, and we have an eigenvalue. Its eigenvector $\boldsymbol{x}$ is positive because on the left side of that equality, $A\boldsymbol{x}$ is sure to be positive.

To see that no eigenvalue can be larger than $t_{\max}$, suppose $A\boldsymbol{z} = \lambda\boldsymbol{z}$. Since λ and $\boldsymbol{z}$ may involve negative or complex numbers, we take absolute values: $|\lambda||\boldsymbol{z}| = |A\boldsymbol{z}| \leq A|\boldsymbol{z}|$ by the "triangle inequality." This $|\boldsymbol{z}|$ is a nonnegative vector, so $|\lambda|$ is one of the possible candidates t. Therefore $|\lambda|$ cannot exceed $t_{\max}$—which must be $\lambda_{\max}$.

Population Growth

Divide the population into three age groups: age < 20, age 20 to 39, and age 40 to 59. At year T the sizes of those groups are n_1, n_2, n_3. Twenty years later, the sizes have changed for two reasons:

1. **Reproduction** $n_1^{\text{new}} = F_1 n_1 + F_2 n_2 + F_3 n_3$ gives a new generation

2. **Survival** $n_2^{\text{new}} = P_1 n_1$ and $n_3^{\text{new}} = P_2 n_2$ gives the older generations

The fertility rates are F_1, F_2, F_3 (F_2 largest). The *Leslie matrix* A might look like this:

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}^{\text{new}} = \begin{bmatrix} F_1 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} .04 & \mathbf{1.1} & .01 \\ .98 & 0 & 0 \\ 0 & \mathbf{.92} & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}.$$

This is population projection in its simplest form, the same matrix A at every step. In a realistic model, A will change with time (from the environment or internal factors). Professors may want to include a fourth group, age ≥ 60, but we don't allow it.

The matrix has $A \geq 0$ but not $A > 0$. The Perron-Frobenius theorem still applies because $A^3 > 0$. The largest eigenvalue is $\lambda_{\max} \approx 1.06$. You can watch the generations move, starting from $n_2 = 1$ in the middle generation:

$$\mathbf{eig}(A) = \begin{matrix} \mathbf{1.06} \\ -1.01 \\ -0.01 \end{matrix} \qquad A^2 = \begin{bmatrix} 1.08 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{1.08} & .01 \\ 0.90 & 0 & 0 \end{bmatrix} \qquad A^3 = \begin{bmatrix} 0.10 & \mathbf{1.19} & .01 \\ 0.06 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{0.99} & .01 \end{bmatrix}.$$

A fast start would come from $\boldsymbol{u}_0 = (0, 1, 0)$. That middle group will reproduce 1.1 and also survive .92. The newest and oldest generations are in $\boldsymbol{u}_1 = (1.1, 0, .92) =$ column 2 of A. Then $\boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$ is the second column of A^2. The early numbers (transients) depend a lot on $\boldsymbol{u}_0$, but ***the asymptotic growth rate*** $\lambda_{\max}$ ***is the same from every start.*** Its eigenvector $\boldsymbol{x} = (.63, .58, .51)$ shows all three groups growing steadily together.

Caswell's book on *Matrix Population Models* emphasizes sensitivity analysis. The model is never exactly right. If the F's or P's in the matrix change by 10%, does $\lambda_{\max}$ go below 1 (which means extinction)? Problem 19 will show that a matrix change ΔA produces an eigenvalue change $\Delta\lambda = \boldsymbol{y}^{\mathrm{T}}(\Delta A)\boldsymbol{x}$. Here $\boldsymbol{x}$ and $\boldsymbol{y}^{\mathrm{T}}$ are the right and left eigenvectors of A. So $\boldsymbol{x}$ is a column of S and $\boldsymbol{y}^{\mathrm{T}}$ is a row of S^{-1}.

Linear Algebra in Economics: The Consumption Matrix

A long essay about linear algebra in economics would be out of place here. A short note about one matrix seems reasonable. The ***consumption matrix*** tells how much of each input goes into a unit of output. This describes the manufacturing side of the economy.

Consumption matrix We have n industries like chemicals, food, and oil. To produce a unit of chemicals may require .2 units of chemicals, .3 units of food, and .4 units of oil. Those numbers go into row 1 of the consumption matrix A:

$$\begin{bmatrix} \text{chemical output} \\ \text{food output} \\ \text{oil output} \end{bmatrix} = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix} \begin{bmatrix} \text{chemical input} \\ \text{food input} \\ \text{oil input} \end{bmatrix}.$$

Row 2 shows the inputs to produce food—a heavy use of chemicals and food, not so much oil. Row 3 of A shows the inputs consumed to refine a unit of oil. The real consumption matrix for the United States in 1958 contained 83 industries. The models in the 1990's are much larger and more precise. We chose a consumption matrix that has a convenient eigenvector.

Now comes the question: Can this economy meet demands y_1, y_2, y_3 for chemicals, food, and oil? To do that, the inputs p_1, p_2, p_3 will have to be higher—because part of $\boldsymbol{p}$ is consumed in producing $\boldsymbol{y}$. The input is $\boldsymbol{p}$ and the consumption is $A\boldsymbol{p}$, which leaves the output $\boldsymbol{p} - A\boldsymbol{p}$. This net production is what meets the demand $\boldsymbol{y}$:

Problem Find a vector $\boldsymbol{p}$ such that $\quad \boldsymbol{p} - A\boldsymbol{p} = \boldsymbol{y} \quad$ or $\quad \boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$.

Apparently the linear algebra question is whether $I - A$ is invertible. But there is more to the problem. The demand vector $\boldsymbol{y}$ is nonnegative, and so is A. *The production levels in* $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$ *must also be nonnegative*. The real question is:

When is $(I - A)^{-1}$ a nonnegative matrix?

This is the test on $(I - A)^{-1}$ for a productive economy, which can meet any positive demand. If A is small compared to I, then $A\boldsymbol{p}$ is small compared to $\boldsymbol{p}$. There is plenty of output. If A is too large, then production consumes more than it yields. In this case the external demand $\boldsymbol{y}$ cannot be met.

"Small" or "large" is decided by the largest eigenvalue λ_1 of A (which is positive):

If $\lambda_1 > 1$ then $(I - A)^{-1}$ has negative entries
If $\lambda_1 = 1$ then $(I - A)^{-1}$ fails to exist
If $\lambda_1 < 1$ then $(I - A)^{-1}$ is nonnegative as desired.

The main point is that last one. The reasoning uses a nice formula for $(I - A)^{-1}$, which we give now. The most important infinite series in mathematics is the ***geometric series*** $1 + x + x^2 + \cdots$. This series adds up to $1/(1 - x)$ provided x lies between -1 and 1. When $x = 1$ the series is $1 + 1 + 1 + \cdots = \infty$. When $|x| \geq 1$ the terms x^n don't go to zero and the series has no chance to converge.

The nice formula for $(I - A)^{-1}$ is the ***geometric series of matrices***:

Geometric series $\quad (I - A)^{-1} = I + A + A^2 + A^3 + \cdots.$

If you multiply the series $S = I + A + A^2 + \cdots$ by A, you get the same series except for I. Therefore $S - AS = I$, which is $(I - A)S = I$. The series adds to $S = (I - A)^{-1}$ if it converges. ***And it converges if all eigenvalues of A have $|\lambda| < 1$.***

In our case $A \geq 0$. All terms of the series are nonnegative. Its sum is $(I - A)^{-1} \geq 0$.

Example 4 $A = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix}$ has $\lambda_{\max} = .9$ and $(I - A)^{-1} = \frac{1}{93}\begin{bmatrix} 41 & 25 & 27 \\ 33 & 36 & 24 \\ 34 & 23 & 36 \end{bmatrix}$.

This economy is productive. A is small compared to I, because $\lambda_{\max}$ is .9. To meet the demand $\boldsymbol{y}$, start from $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$. Then $A\boldsymbol{p}$ is consumed in production, leaving $\boldsymbol{p} - A\boldsymbol{p}$. This is $(I - A)\boldsymbol{p} = \boldsymbol{y}$, and the demand is met.

Example 5 $A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$ has $\lambda_{\max} = 2$ and $(I - A)^{-1} = -\frac{1}{3}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$.

This consumption matrix A is too large. Demands can't be met, because production consumes more than it yields. The series $I + A + A^2 + \ldots$ does not converge to $(I - A)^{-1}$ because $\lambda_{\max} > 1$. The series is growing while $(I - A)^{-1}$ is actually negative.

In the same way $1 + 2 + 4 + \cdots$ is not really $1/(1 - 2) = -1$. But not entirely false!

Problem Set 8.3

Questions 1–12 are about Markov matrices and their eigenvalues and powers.

1 Find the eigenvalues of this Markov matrix (their sum is the trace):

$$A = \begin{bmatrix} .90 & .15 \\ .10 & .85 \end{bmatrix}.$$

What is the steady state eigenvector for the eigenvalue $\lambda_1 = 1$?

2 Diagonalize the Markov matrix in Problem 1 to $A = S\Lambda S^{-1}$ by finding its other eigenvector:

$$A = \begin{bmatrix} & \\ & \end{bmatrix}\begin{bmatrix} 1 & \\ & .75 \end{bmatrix}\begin{bmatrix} & \\ & \end{bmatrix}.$$

What is the limit of $A^k = S\Lambda^k S^{-1}$ when $\Lambda^k = \begin{bmatrix} 1 & 0 \\ 0 & .75^k \end{bmatrix}$ approaches $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$?

3 What are the eigenvalues and steady state eigenvectors for these Markov matrices?

$$A = \begin{bmatrix} 1 & .2 \\ 0 & .8 \end{bmatrix} \quad A = \begin{bmatrix} .2 & 1 \\ .8 & 0 \end{bmatrix} \quad A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

4 For every 4 by 4 Markov matrix, what eigenvector of A^{T} corresponds to the (known) eigenvalue $\lambda = 1$?

5 Every year 2% of young people become old and 3% of old people become dead. (No births.) Find the steady state for

$$\begin{bmatrix} \text{young} \\ \text{old} \\ \text{dead} \end{bmatrix}_{k+1} = \begin{bmatrix} .98 & .00 & 0 \\ .02 & .97 & 0 \\ .00 & .03 & 1 \end{bmatrix} \begin{bmatrix} \text{young} \\ \text{old} \\ \text{dead} \end{bmatrix}_k .$$

6 For a Markov matrix, the sum of the components of $\boldsymbol{x}$ equals the sum of the components of $A\boldsymbol{x}$. If $A\boldsymbol{x} = \lambda\boldsymbol{x}$ with $\lambda \neq 1$, prove that the components of this non-steady eigenvector $\boldsymbol{x}$ add to zero.

7 Find the eigenvalues and eigenvectors of A. Explain why A^k approaches A^∞:

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \qquad A^\infty = \begin{bmatrix} .6 & .6 \\ .4 & .4 \end{bmatrix}.$$

Challenge problem: Which Markov matrices produce that steady state (.6, .4)?

8 The steady state eigenvector of a permutation matrix is $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. This is *not* approached when $\boldsymbol{u}_0 = (0, 0, 0, 1)$. What are $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ and $\boldsymbol{u}_4$? What are the four eigenvalues of P, which solve $\lambda^4 = 1$?

Permutation matrix = Markov matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

9 Prove that the square of a Markov matrix is also a Markov matrix.

10 If $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is a Markov matrix, its eigenvalues are 1 and ______. The steady state eigenvector is $\boldsymbol{x}_1 =$ ______.

11 Complete A to a Markov matrix and find the steady state eigenvector. When A is a symmetric Markov matrix, why is $\boldsymbol{x}_1 = (1, \ldots, 1)$ its steady state?

$$A = \begin{bmatrix} .7 & .1 & .2 \\ .1 & .6 & .3 \\ \_ & \_ & \_ \end{bmatrix}.$$

12 A Markov differential equation is not $d\boldsymbol{u}/dt = A\boldsymbol{u}$ but $d\boldsymbol{u}/dt = (A - I)\boldsymbol{u}$. The diagonal is negative, the rest of $A - I$ is positive. The columns add to zero.

Find the eigenvalues of $B = A - I = \begin{bmatrix} -.2 & .3 \\ .2 & -.3 \end{bmatrix}$. Why does $A - I$ have $\lambda = 0$?

When $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$ multiply $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, what is the steady state as $t \to \infty$?

Questions 13–15 are about linear algebra in economics.

13 Each row of the consumption matrix in Example 4 adds to .9. Why does that make $\lambda = .9$ an eigenvalue, and what is the eigenvector?

14 Multiply $I + A + A^2 + A^3 + \cdots$ by $I - A$ to show that the series adds to _____. For $A = \left[\begin{smallmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{smallmatrix}\right]$, find A^2 and A^3 and use the pattern to add up the series.

15 For which of these matrices does $I + A + A^2 + \cdots$ yield a nonnegative matrix $(I - A)^{-1}$? Then the economy can meet any demand:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 4 \\ .2 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} .5 & 1 \\ .5 & 0 \end{bmatrix}.$$

If the demands are $\boldsymbol{y} = (2, 6)$, what are the vectors $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$?

16 (Markov again) This matrix has zero determinant. What are its eigenvalues?

$$A = \begin{bmatrix} .4 & .2 & .3 \\ .2 & .4 & .3 \\ .4 & .4 & .4 \end{bmatrix}.$$

Find the limits of $A^k\boldsymbol{u}_0$ starting from $\boldsymbol{u}_0 = (1, 0, 0)$ and then $\boldsymbol{u}_0 = (100, 0, 0)$.

17 If A is a Markov matrix, does $I + A + A^2 + \cdots$ add up to $(I - A)^{-1}$?

18 For the Leslie matrix show that $\det(A - \lambda I) = 0$ gives $F_1\lambda^2 + F_2P_1\lambda + F_3P_1P_2 = \lambda^3$. The right side λ^3 is larger as $\lambda \longrightarrow \infty$. The left side is larger at $\lambda = 1$ if $F_1 + F_2P_1 + F_3P_1P_2 > 1$. In that case the two sides are equal at an eigenvalue $\lambda_{\max} > 1$: *growth.*

19 **Sensitivity of eigenvalues**: A matrix change ΔA produces eigenvalue changes $\Delta\Lambda$. *The formula for those changes* $\Delta\lambda_1, \ldots, \Delta\lambda_n$ *is* **diag**$(S^{-1}\,\Delta A\,S)$. **Challenge**: Start from $AS = S\Lambda$. The eigenvectors and eigenvalues change by ΔS and $\Delta\Lambda$:

$$(A{+}\Delta A)(S{+}\Delta S) = (S{+}\Delta S)(\Lambda{+}\Delta\Lambda) \text{ becomes } A(\Delta S){+}(\Delta A)S = S(\Delta\Lambda){+}(\Delta S)\Lambda.$$

Small terms $(\Delta A)(\Delta S)$ and $(\Delta S)(\Delta\Lambda)$ are ignored. *Multiply the last equation by* S^{-1}. From the inner terms, the diagonal part of $S^{-1}(\Delta A)S$ gives $\Delta\Lambda$ as we want. *Why do the outer terms* $S^{-1}\,A\,\Delta S$ *and* $S^{-1}\,\Delta S\,\Lambda$ *cancel on the diagonal?*

$$\text{Explain } S^{-1}A = \Lambda S^{-1} \text{ and then } \quad \mathbf{diag}(\Lambda\,S^{-1}\,\Delta S) = \mathbf{diag}(S^{-1}\,\Delta S\,\Lambda).$$

20 Suppose $B > A > 0$, meaning that each $b_{ij} > a_{ij} > 0$. How does the Perron-Frobenius discussion show that $\lambda_{\max}(B) > \lambda_{\max}(A)$?