Basic Linear Algebra

In this chapter, we study the topics in linear algebra that will be needed in the rest of the book. We begin by discussing the building blocks of linear algebra: matrices and vectors. Then we use our knowledge of matrices and vectors to develop a systematic procedure (the Gauss–Jordan method) for solving linear equations, which we then use to invert matrices. We close the chapter with an introduction to determinants.

The material covered in this chapter will be used in our study of linear and nonlinear programming.

2.1 Matrices and Vectors

Matrices

DEFINITION ■ A **matrix** is any rectangular array of numbers. ■

For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \end{bmatrix}$$

are all matrices.

If a matrix A has m rows and n columns, we call A an $m \times n$ matrix. We refer to $m \times n$ as the **order** of the matrix. A typical $m \times n$ matrix A may be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

DEFINITION ■ The number in the ith row and jth column of A is called the ***ij*th element** of A and is written a_{ij}. ■

For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

then $a_{11} = 1$, $a_{23} = 6$, and $a_{31} = 7$.

Sometimes we will use the notation $A = [a_{ij}]$ to indicate that A is the matrix whose ijth element is a_{ij}.

DEFINITION ■ Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are **equal** if and only if A and B are of the same order and for all i and j, $a_{ij} = b_{ij}$. ■

For example, if

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} x & y \\ w & z \end{bmatrix}$$

then $A = B$ if and only if $x = 1$, $y = 2$, $w = 3$, and $z = 4$.

Vectors

Any matrix with only one column (that is, any $m \times 1$ matrix) may be thought of as a **column vector.** The number of rows in a column vector is the **dimension** of the column vector. Thus,

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

may be thought of as a 2×1 matrix or a two-dimensional column vector. R^m will denote the set of all m-dimensional column vectors.

In analogous fashion, we can think of any vector with only one row (a $1 \times n$ matrix as a **row vector.** The dimension of a row vector is the number of columns in the vector. Thus, $[9 \quad 2 \quad 3]$ may be viewed as a 1×3 matrix or a three-dimensional row vector. In this book, vectors appear in boldface type: for instance, vector **v.** An m-dimensional vector (either row or column) in which all elements equal zero is called a **zero vector** (written **0**). Thus,

$$[0 \quad 0] \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

are two-dimensional zero vectors.

Any m-dimensional vector corresponds to a directed line segment in the m-dimensional plane. For example, in the two-dimensional plane, the vector

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

corresponds to the line segment joining the point

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

to the point

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The directed line segments corresponding to

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

are drawn in Figure 1.

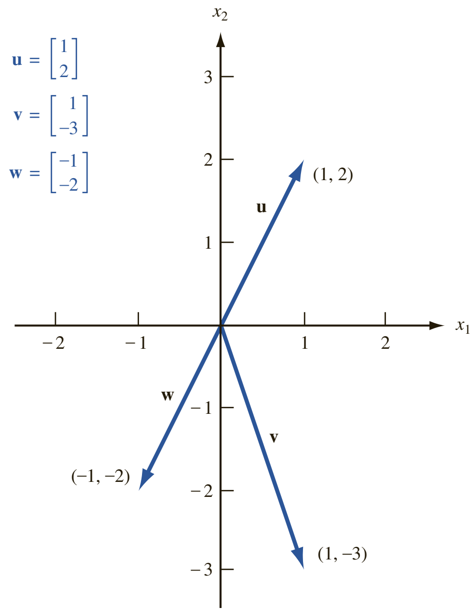


FIGURE 1
Vectors Are Directed Line Segments

The Scalar Product of Two Vectors

An important result of multiplying two vectors is the *scalar product.* To define the scalar product of two vectors, suppose we have a row vector $\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n]$ and a column vector

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

of the same dimension. The **scalar product** of $\mathbf{u}$ and $\mathbf{v}$ (written $\mathbf{u} \cdot \mathbf{v}$) is the number $u_1v_1 + u_2v_2 + \cdots + u_nv_n$.

For the scalar product of two vectors to be defined, the first vector must be a row vector and the second vector must be a column vector. For example, if

$$\mathbf{u} = [1 \quad 2 \quad 3] \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

then $\mathbf{u} \cdot \mathbf{v} = 1(2) + 2(1) + 3(2) = 10$. By these rules for computing a scalar product, if

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{v} = [2 \quad 3]$$

then $\mathbf{u} \cdot \mathbf{v}$ is not defined. Also, if

$$\mathbf{u} = [1 \quad 2 \quad 3] \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

then $\mathbf{u} \cdot \mathbf{v}$ is not defined because the vectors are of two different dimensions.

Note that two vectors are perpendicular if and only if their scalar product equals 0. Thus, the vectors $[1 \quad -1]$ and $[1 \quad 1]$ are perpendicular.

We note that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos \theta$, where $\|\mathbf{u}\|$ is the length of the vector $\mathbf{u}$ and θ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$.

Matrix Operations

We now describe the arithmetic operations on matrices that are used later in this book.

The Scalar Multiple of a Matrix

Given any matrix A and any number c (a *number* is sometimes referred to as a *scalar*), the matrix cA is obtained from the matrix A by multiplying each element of A by c. For example,

$$\text{if} \qquad A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}, \qquad \text{then} \qquad 3A = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$$

For $c = -1$, scalar multiplication of the matrix A is sometimes written as $-A$.

Addition of Two Matrices

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices with the same order (say, $m \times n$). Then the matrix $C = A + B$ is defined to be the $m \times n$ matrix whose ijth element is $a_{ij} + b_{ij}$. Thus, to obtain the sum of two matrices A and B, we add the corresponding elements of A and B. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 1 & -1 \end{bmatrix}$$

then

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

This rule for matrix addition may be used to add vectors of the same dimension. For example, if $\mathbf{u} = [1 \quad 2]$ and $\mathbf{v} = [2 \quad 1]$, then $\mathbf{u} + \mathbf{v} = [1+2 \quad 2+1] = [3 \quad 3]$. Vectors may be added geometrically by the parallelogram law (see Figure 2).

We can use scalar multiplication and the addition of matrices to define the concept of a line segment. A glance at Figure 1 should convince you that any point u in the m-dimensional plane corresponds to the m-dimensional vector $\mathbf{u}$ formed by joining the origin to the point u. For any two points u and v in the m-dimensional plane, the **line segment** joining u and v (called the line segment uv) is the set of all points in the m-dimensional plane that correspond to the vectors $c\mathbf{u} + (1 - c)\mathbf{v}$, where $0 \le c \le 1$ (Figure 3). For example, if $u = (1, 2)$ and $v = (2, 1)$, then the line segment uv consists

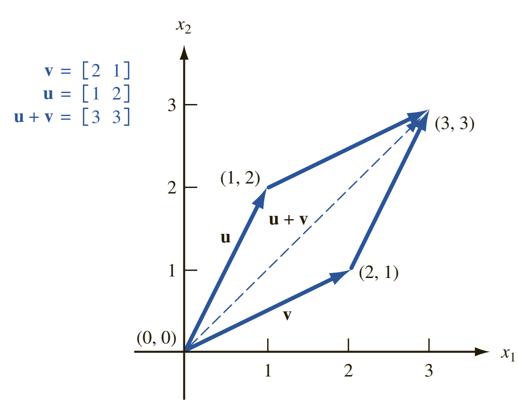


FIGURE 2
Addition of Vectors

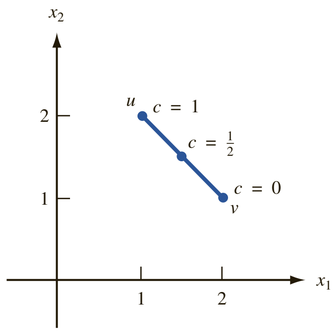


FIGURE 3
Line Segment Joining $u = (1, 2)$ and $v = (2, 1)$

of the points corresponding to the vectors $c[1 \quad 2] + (1 - c)[2 \quad 1] = [2 - c \quad 1 + c]$, where $0 \leq c \leq 1$. For $c = 0$ and $c = 1$, we obtain the endpoints of the line segment uv; for $c = \frac{1}{2}$, we obtain the midpoint $(0.5\mathbf{u} + 0.5\mathbf{v})$ of the line segment uv.

Using the parallelogram law, the line segment uv may also be viewed as the points corresponding to the vectors $\mathbf{u} + c(\mathbf{v} - \mathbf{u})$, where $0 \leq c \leq 1$ (Figure 4). Observe that for $c = 0$, we obtain the vector $\mathbf{u}$ (corresponding to point u), and for $c = 1$, we obtain the vector $\mathbf{v}$ (corresponding to point v).

The Transpose of a Matrix

Given any $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

the **transpose** of A (written A^T) is the $n \times m$ matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

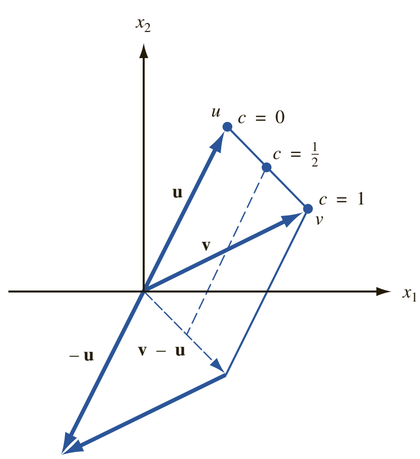


FIGURE 4
Representation of Line Segment uv

Thus, A^T is obtained from A by letting row 1 of A be column 1 of A^T, letting row 2 of A be column 2 of A^T, and so on. For example,

$$\text{if} \qquad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad \text{then} \qquad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Observe that $(A^T)^T = A$. Let $B = [1 \quad 2]$; then

$$B^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad (B^T)^T = [1 \quad 2] = B$$

As indicated by these two examples, for any matrix A, $(A^T)^T = A$.

Matrix Multiplication

Given two matrices A and B, the matrix product of A and B (written AB) is defined if and only if

$$\text{Number of columns in } A = \text{number of rows in } B \tag{1}$$

For the moment, assume that for some positive integer r, A has r columns and B has r rows. Then for some m and n, A is an $m \times r$ matrix and B is an $r \times n$ matrix.

DEFINITION ■ The **matrix product** $C = AB$ of A and B is the $m \times n$ matrix C whose ijth element is determined as follows:

$$ij\text{th element of } C = \text{scalar product of row } i \text{ of } A \times \text{column } j \text{ of } B \quad ■ \tag{2}$$

If Equation (1) is satisfied, then each row of A and each column of B will have the same number of elements. Also, if (1) is satisfied, then the scalar product in Equation (2) will be defined. The product matrix $C = AB$ will have the same number of rows as A and the same number of columns as B.

EXAMPLE 1 Matrix Multiplication

Compute $C = AB$ for

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}$$

Solution Because A is a 2×3 matrix and B is a 3×2 matrix, AB is defined, and C will be a 2×2 matrix. From Equation (2),

$$c_{11} = [1 \quad 1 \quad 2] \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 1(1) + 1(2) + 2(1) = 5$$

$$c_{12} = [1 \quad 1 \quad 2] \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 1(1) + 1(3) + 2(2) = 8$$

$$c_{21} = [2 \quad 1 \quad 3] \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 2(1) + 1(2) + 3(1) = 7$$

$$c_{22} = [2 \quad 1 \quad 3]\begin{bmatrix}1\\3\\2\end{bmatrix} = 2(1) + 1(3) + 3(2) = 11$$

$$C = AB = \begin{bmatrix}5 & 8\\7 & 11\end{bmatrix}$$

EXAMPLE 2 Column Vector Times Row Vector

Find AB for

$$A = \begin{bmatrix}3\\4\end{bmatrix} \quad \text{and} \quad B = [1 \quad 2]$$

Solution Because A has one column and B has one row, $C = AB$ will exist. From Equation (2), we know that C is a 2×2 matrix with

$$c_{11} = 3(1) = 3 \qquad c_{21} = 4(1) = 4$$
$$c_{12} = 3(2) = 6 \qquad c_{22} = 4(2) = 8$$

Thus,

$$C = \begin{bmatrix}3 & 6\\4 & 8\end{bmatrix}$$

EXAMPLE 3 Row Vector Times Column Vector

Compute $D = BA$ for the A and B of Example 2.

Solution In this case, D will be a 1×1 matrix (or a scalar). From Equation (2),

$$d_{11} = [1 \quad 2]\begin{bmatrix}3\\4\end{bmatrix} = 1(3) + 2(4) = 11$$

Thus, $D = [11]$. In this example, matrix multiplication is equivalent to scalar multiplication of a row and column vector.

Recall that if you multiply two real numbers a and b, then $ab = ba$. This is called the *commutative property of multiplication.* Examples 2 and 3 show that for matrix multiplication, it may be that $AB \neq BA$. Matrix multiplication is not necessarily commutative. (In some cases, however, $AB = BA$ will hold.)

EXAMPLE 4 Undefined Matrix Product

Show that AB is undefined if

$$A = \begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix}1 & 1\\0 & 1\\1 & 2\end{bmatrix}$$

Solution This follows because A has two columns and B has three rows. Thus, Equation (1) is not satisfied.

TABLE 1
Gallons of Crude Oil Required to Produce 1 Gallon of Gasoline

Crude Oil	Premium Unleaded	Regular Unleaded	Regular Leaded
1	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$

Many computations that commonly occur in operations research (and other branches of mathematics) can be concisely expressed by using matrix multiplication. To illustrate this, suppose an oil company manufactures three types of gasoline: premium unleaded, regular unleaded, and regular leaded. These gasolines are produced by mixing two types of crude oil: crude oil 1 and crude oil 2. The number of gallons of crude oil required to manufacture 1 gallon of gasoline is given in Table 1.

From this information, we can find the amount of each type of crude oil needed to manufacture a given amount of gasoline. For example, if the company wants to produce 10 gallons of premium unleaded, 6 gallons of regular unleaded, and 5 gallons of regular leaded, then the company's crude oil requirements would be

$$\text{Crude 1 required} = (\tfrac{3}{4})(10) + (\tfrac{2}{3})(6) + (\tfrac{1}{4})5 = 12.75 \text{ gallons}$$
$$\text{Crude 2 required} = (\tfrac{1}{4})(10) + (\tfrac{1}{3})(6) + (\tfrac{3}{4})5 = 8.25 \text{ gallons}$$

More generally, we define

$$\begin{aligned}
p_U &= \text{gallons of premium unleaded produced}\\
r_U &= \text{gallons of regular unleaded produced}\\
r_L &= \text{gallons of regular leaded produced}\\
c_1 &= \text{gallons of crude 1 required}\\
c_2 &= \text{gallons of crude 2 required}
\end{aligned}$$

Then the relationship between these variables may be expressed by

$$\begin{aligned}
c_1 &= (\tfrac{3}{4})\,p_U + (\tfrac{2}{3})\,r_U + (\tfrac{1}{4})\,r_L\\
c_2 &= (\tfrac{1}{4})\,p_U + (\tfrac{1}{3})\,r_U + (\tfrac{3}{4})\,r_L
\end{aligned}$$

Using matrix multiplication, these relationships may be expressed by

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{2}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{3} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} p_U \\ r_U \\ r_L \end{bmatrix}$$

Properties of Matrix Multiplication

To close this section, we discuss some important properties of matrix multiplication. In what follows, we assume that all matrix products are defined.

1 Row i of AB = (row i of A)B. To illustrate this property, let

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}$$

Then row 2 of the 2×2 matrix AB is equal to

$$[2 \quad 1 \quad 3]\begin{bmatrix}1 & 1\\ 2 & 3\\ 1 & 2\end{bmatrix} = [7 \quad 11]$$

This answer agrees with Example 1.

2 Column j of $AB = A$(column j of B). Thus, for A and B as given, the first column of AB is

$$\begin{bmatrix}1 & 1 & 2\\ 2 & 1 & 3\end{bmatrix}\begin{bmatrix}1\\ 2\\ 1\end{bmatrix} = \begin{bmatrix}5\\ 7\end{bmatrix}$$

Properties 1 and 2 are helpful when you need to compute only *part* of the matrix AB.

3 Matrix multiplication is associative. That is, $A(BC) = (AB)C$. To illustrate, let

$$A = [1 \quad 2], \qquad B = \begin{bmatrix}2 & 3\\ 4 & 5\end{bmatrix}, \qquad C = \begin{bmatrix}2\\ 1\end{bmatrix}$$

Then $AB = [10 \quad 13]$ and $(AB)C = 10(2) + 13(1) = [33]$.

On the other hand,

$$BC = \begin{bmatrix}7\\ 13\end{bmatrix}$$

so $A(BC) = 1(7) + 2(13) = [33]$. In this case, $A(BC) = (AB)C$ does hold.

4 Matrix multiplication is distributive. That is, $A(B + C) = AB + AC$ and $(B + C)D = BD + CD$.

Matrix Multiplication with Excel

Using the Excel MMULT function, it is easy to multiply matrices. To illustrate, let's use Excel to find the matrix product AB that we found in Example 1 (see Figure 5 and file Mmult.xls). We proceed as follows:

Mmult.xls

Step 1 Enter A and B in D2:F3 and D5:E7, respectively.

Step 2 Select the range (D9:E10) in which the product AB will be computed.

Step 3 In the upper left-hand corner (D9) of the selected range, type the formula

$$= \text{MMULT(D2:F3,D5:E7)}$$

Then hit **Control Shift Enter** (not just Enter), and the desired matrix product will be computed. Note that MMULT is an *array* function and not an ordinary spreadsheet function. This explains why we must preselect the range for AB and use Control Shift Enter.

	A	B	C	D	E	F
1	MatrixMultiplication					
2				1	1	2
3			A	2	1	3
4						
5			B	1	1	
6				2	3	
7				1	2	
8						
9				5	8	
10			C	7	11	
11						

FIGURE 5

P R O B L E M S

Group A

1 For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 1 & 2 \end{bmatrix}$, find:

a $-A$ **b** $3A$ **c** $A + 2B$
d A^T **e** B^T **f** AB
g BA

2 Only three brands of beer (beer 1, beer 2, and beer 3) are available for sale in Metropolis. From time to time, people try one or another of these brands. Suppose that at the beginning of each month, people change the beer they are drinking according to the following rules:

30% of the people who prefer beer 1 switch to beer 2.
20% of the people who prefer beer 1 switch to beer 3.
30% of the people who prefer beer 2 switch to beer 3.
30% of the people who prefer beer 3 switch to beer 2.
10% of the people who prefer beer 3 switch to beer 1.

For $i = 1, 2, 3$, let x_i be the number who prefer beer i at the beginning of this month and y_i be the number who prefer beer i at the beginning of next month. Use matrix multiplication to relate the following:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Group B

3 Prove that matrix multiplication is associative.

4 Show that for any two matrices A and B, $(AB)^T = B^T A^T$.

5 An $n \times n$ matrix A is symmetric if $A = A^T$.

a Show that for any $n \times n$ matrix, AA^T is a symmetric matrix.

b Show that for any $n \times n$ matrix A, $(A + A^T)$ is a symmetric matrix.

6 Suppose that A and B are both $n \times n$ matrices. Show that computing the matrix product AB requires n^3 multiplications and $n^3 - n^2$ additions.

7 The **trace of a matrix** is the sum of its diagonal elements.

a For any two matrices A and B, show that trace $(A + B) =$ trace $A +$ trace B.

b For any two matrices A and B for which the products AB and BA are defined, show that trace $AB =$ trace BA.

2.2 Matrices and Systems of Linear Equations

Consider a system of linear equations given by

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad\qquad \vdots \quad & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{3}$$

In Equation (3), $x_1, x_2, \ldots, x_n$ are referred to as **variables,** or unknowns, and the a_{ij}'s and b_i's are **constants.** A set of equations such as (3) is called a linear system of m equations in n variables.

DEFINITION ■ A **solution** to a linear system of m equations in n unknowns is a set of values for the unknowns that satisfies each of the system's m equations. ■

To understand linear programming, we need to know a great deal about the properties of solutions to linear equation systems. With this in mind, we will devote much effort to studying such systems.

We denote a possible solution to Equation (3) by an n-dimensional column vector **x,** in which the ith element of **x** is the value of x_i. The following example illustrates the concept of a solution to a linear system.

EXAMPLE 5 Solution to Linear System

Show that

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is a solution to the linear system

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ 2x_1 - x_2 &= 0 \end{aligned} \tag{4}$$

and that

$$\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

is not a solution to linear system (4).

Solution To show that

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is a solution to Equation (4), we substitute $x_1 = 1$ and $x_2 = 2$ in both equations and check that they are satisfied: $1 + 2(2) = 5$ and $2(1) - 2 = 0$.

The vector

$$\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

is not a solution to (4), because $x_1 = 3$ and $x_2 = 1$ fail to satisfy $2x_1 - x_2 = 0$.

Using matrices can greatly simplify the statement and solution of a system of linear equations. To show how matrices can be used to compactly represent Equation (3), let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Then (3) may be written as

$$A\mathbf{x} = \mathbf{b} \tag{5}$$

Observe that both sides of Equation (5) will be $m \times 1$ matrices (or $m \times 1$ column vectors). For the matrix $A\mathbf{x}$ to equal the matrix $\mathbf{b}$ (or for the vector $A\mathbf{x}$ to equal the vector $\mathbf{b}$), their corresponding elements must be equal. The first element of $A\mathbf{x}$ is the scalar product of row 1 of A with $\mathbf{x}$. This may be written as

$$[a_{11} \quad a_{12} \quad \cdots \quad a_{1n}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$$

This must equal the first element of $\mathbf{b}$ (which is b_1). Thus, (5) implies that $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$. This is the first equation of (3). Similarly, (5) implies that the scalar product of row i of A with $\mathbf{x}$ must equal b_i, and this is just the ith equation of (3). Our discussion shows that (3) and (5) are two different ways of writing the same linear system. We call (5) the **matrix representation** of (3). For example, the matrix representation of (4) is

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

Sometimes we abbreviate (5) by writing

$$A|\mathbf{b} \tag{6}$$

If A is an $m \times n$ matrix, it is assumed that the variables in (6) are $x_1, x_2, \ldots, x_n$. Then (6) is still another representation of (3). For instance, the matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{array}\right]$$

represents the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 2 \\ x_2 + 2x_3 &= 3 \\ x_1 + x_2 + x_3 &= 1 \end{aligned}$$

PROBLEM

Group A

1 Use matrices to represent the following system of equations in two different ways:

$$\begin{aligned} x_1 - x_2 \quad &= 4 \\ 2x_1 + x_2 \quad &= 6 \\ x_1 + 3x_2 &= 8 \end{aligned}$$

2.3 The Gauss–Jordan Method for Solving Systems of Linear Equations

We develop in this section an efficient method (the Gauss–Jordan method) for solving a system of linear equations. Using the Gauss–Jordan method, we show that any system of linear equations must satisfy one of the following three cases:

Case 1 The system has no solution.

Case 2 The system has a unique solution.

Case 3 The system has an infinite number of solutions.

The Gauss–Jordan method is also important because many of the manipulations used in this method are used when solving linear programming problems by the simplex algorithm (see Chapter 4).

Elementary Row Operations

Before studying the Gauss–Jordan method, we need to define the concept of an **elementary row operation** (ERO). An ERO transforms a given matrix A into a new matrix A' via one of the following operations.

Type 1 ERO

A' is obtained by multiplying any row of A by a nonzero scalar. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 6 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

then a Type 1 ERO that multiplies row 2 of A by 3 would yield

$$A' = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 9 & 15 & 18 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Type 2 ERO

Begin by multiplying any row of A (say, row i) by a nonzero scalar c. For some $j \neq i$, let row j of $A' = c$(row i of A) + row j of A, and let the other rows of A' be the same as the rows of A.

For example, we might multiply row 2 of A by 4 and replace row 3 of A by 4(row 2 of A) + row 3 of A. Then row 3 of A' becomes

$$4\,[1 \quad 3 \quad 5 \quad 6] + [0 \quad 1 \quad 2 \quad 3] = [4 \quad 13 \quad 22 \quad 27]$$

and

$$A' = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 6 \\ 4 & 13 & 22 & 27 \end{bmatrix}$$

Type 3 ERO

Interchange any two rows of A. For instance, if we interchange rows 1 and 3 of A, we obtain

$$A' = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 3 & 5 & 6 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

Type 1 and Type 2 EROs formalize the operations used to solve a linear equation system. To solve the system of equations

$$\begin{aligned} x_1 + x_2 &= 2 \\ 2x_1 + 4x_2 &= 7 \end{aligned} \tag{7}$$

we might proceed as follows. First replace the second equation in (7) by -2(first equation in (7)) + second equation in (7). This yields the following linear system:

$$\begin{aligned} x_1 + x_2 &= 2 \\ 2x_2 &= 3 \end{aligned} \tag{7.1}$$

Then multiply the second equation in (7.1) by $\frac{1}{2}$, yielding the system

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_2 &= \tfrac{3}{2} \end{aligned} \tag{7.2}$$

Finally, replace the first equation in (7.2) by -1[second equation in (7.2)] + first equation in (7.2). This yields the system

$$\begin{aligned} x_1 \quad &= \tfrac{1}{2} \\ x_2 &= \tfrac{3}{2} \end{aligned} \tag{7.3}$$

System (7.3) has the unique solution $x_1 = \frac{1}{2}$ and $x_2 = \frac{3}{2}$. The systems (7), (7.1), (7.2), and (7.3) are *equivalent* in that they have the same set of solutions. This means that $x_1 = \frac{1}{2}$ and $x_2 = \frac{3}{2}$ is also the unique solution to the original system, (7).

If we view (7) in the augmented matrix form $(A|\mathbf{b})$, we see that the steps used to solve (7) may be seen as Type 1 and Type 2 EROs applied to $A|\mathbf{b}$. Begin with the augmented matrix version of (7):

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 4 & 7 \end{array}\right] \tag{7$'$}$$

Now perform a Type 2 ERO by replacing row 2 of (7′) by −2(row 1 of (7′)) + row 2 of (7′). The result is

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 2 & 3 \end{array}\right] \tag{7.1$'$}$$

which corresponds to (7.1). Next, we multiply row 2 of (7.1′) by $\frac{1}{2}$ (a Type 1 ERO), resulting in

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & \frac{3}{2} \end{array}\right] \tag{7.2$'$}$$

which corresponds to (7.2). Finally, perform a Type 2 ERO by replacing row 1 of (7.2′) by −1(row 2 of (7.2′)) + row 1 of (7.2′). The result is

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{3}{2} \end{array}\right] \tag{7.3$'$}$$

which corresponds to (7.3). Translating (7.3′) back into a linear system, we obtain the system $x_1 = \frac{1}{2}$ and $x_2 = \frac{3}{2}$, which is identical to (7.3).

Finding a Solution by the Gauss–Jordan Method

The discussion in the previous section indicates that if the matrix $A'|\mathbf{b}'$ is obtained from $A|\mathbf{b}$ via an ERO, the systems $A\mathbf{x} = \mathbf{b}$ and $A'\mathbf{x} = \mathbf{b}'$ are equivalent. Thus, any sequence of EROs performed on the augmented matrix $A|\mathbf{b}$ corresponding to the system $A\mathbf{x} = \mathbf{b}$ will yield an equivalent linear system.

The Gauss–Jordan method solves a linear equation system by utilizing EROs in a systematic fashion. We illustrate the method by finding the solution to the following linear system:

$$\begin{aligned} 2x_1 + 2x_2 + x_3 &= 9 \\ 2x_1 - x_2 + 2x_3 &= 6 \\ x_1 - x_2 + 2x_3 &= 5 \end{aligned} \tag{8}$$

The augmented matrix representation is

$$A|\mathbf{b} = \left[\begin{array}{ccc|c} 2 & 2 & 1 & 9 \\ 2 & -1 & 2 & 6 \\ 1 & -1 & 2 & 5 \end{array}\right] \tag{8$'$}$$

Suppose that by performing a sequence of EROs on (8′) we could transform (8′) into

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right] \tag{9'}$$

We note that the result obtained by performing an ERO on a system of equations can also be obtained by multiplying both sides of the matrix representation of the system of equations by a particular matrix. This explains why EROs do not change the set of solutions to a system of equations.

Matrix (9′) corresponds to the following linear system:

$$\begin{aligned} x_1 \qquad &= 1 \\ x_2 \quad &= 2 \\ x_3 &= 3 \end{aligned} \tag{9}$$

System (9) has the unique solution $x_1 = 1, x_2 = 2, x_3 = 3$. Because (9′) was obtained from (8′) by a sequence of EROs, we know that (8) and (9) are equivalent linear systems. Thus, $x_1 = 1, x_2 = 2, x_3 = 3$ must also be the unique solution to (8). We now show how we can use EROs to transform a relatively complicated system such as (8) into a relatively simple system like (9). This is the essence of the Gauss–Jordan method.

We begin by using EROs to transform the first column of (8′) into

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

Then we use EROs to transform the second column of the resulting matrix into

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

Finally, we use EROs to transform the third column of the resulting matrix into

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

As a final result, we will have obtained (9′). We now use the Gauss–Jordan method to solve (8). We begin by using a Type 1 ERO to change the element of (8′) in the first row and first column into a 1. Then we add multiples of row 1 to row 2 and then to row 3 (these are Type 2 EROs). The purpose of these Type 2 EROs is to put zeros in the rest of the first column. The following sequence of EROs will accomplish these goals.

Step 1 Multiply row 1 of (8′) by $\frac{1}{2}$. This Type 1 ERO yields

$$A_1|\mathbf{b}_1 = \left[\begin{array}{ccc|c} 1 & 1 & \frac{1}{2} & \frac{9}{2} \\ 2 & -1 & 2 & 6 \\ 1 & -1 & 2 & 5 \end{array}\right]$$

Step 2 Replace row 2 of $A_1|\mathbf{b}_1$ by -2(row 1 of $A_1|\mathbf{b}_1$) + row 2 of $A_1|\mathbf{b}_1$. The result of this Type 2 ERO is

$$A_2|\mathbf{b}_2 = \left[\begin{array}{ccc|c} 1 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & -3 & 1 & -3 \\ 1 & -1 & 2 & 5 \end{array}\right]$$

Step 3 Replace row 3 of $A_2|\mathbf{b}_2$ by -1(row 1 of $A_2|\mathbf{b}_2$ + row 3 of $A_2|\mathbf{b}_2$. The result of this Type 2 ERO is

$$A_3|\mathbf{b}_3 = \left[\begin{array}{ccc|c} 1 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & -3 & 1 & -3 \\ 0 & -2 & \frac{3}{2} & \frac{1}{2} \end{array}\right]$$

The first column of (8′) has now been transformed into

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

By our procedure, we have made sure that the variable x_1 occurs in only a single equation and in that equation has a coefficient of 1. We now transform the second column of $A_3|\mathbf{b}_3$ into

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

We begin by using a Type 1 ERO to create a 1 in row 2 and column 2 of $A_3|\mathbf{b}_3$. Then we use the resulting row 2 to perform the Type 2 EROs that are needed to put zeros in the rest of column 2. Steps 4–6 accomplish these goals.

Step 4 Multiply row 2 of $A_3|\mathbf{b}_3$ by $-\frac{1}{3}$.The result of this Type 1 ERO is

$$A_4|\mathbf{b}_4 = \left[\begin{array}{ccc|c} 1 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & 1 & -\frac{1}{3} & 1 \\ 0 & -2 & \frac{3}{2} & \frac{1}{2} \end{array}\right]$$

Step 5 Replace row 1 of $A_4|\mathbf{b}_4$ by -1(row 2 of $A_4|\mathbf{b}_4$) + row 1 of $A_4|\mathbf{b}_4$. The result of this Type 2 ERO is

$$A_5|\mathbf{b}_5 = \left[\begin{array}{ccc|c} 1 & 0 & \frac{5}{6} & \frac{7}{2} \\ 0 & 1 & -\frac{1}{3} & 1 \\ 0 & -2 & \frac{3}{2} & \frac{1}{2} \end{array}\right]$$

Step 6 Replace row 3 of $A_5|\mathbf{b}_5$ by 2(row 2 of $A_5|\mathbf{b}_5$) + row 3 of $A_5|\mathbf{b}_5$. The result of this Type 2 ERO is

$$A_6|\mathbf{b}_6 = \left[\begin{array}{ccc|c} 1 & 0 & \frac{5}{6} & \frac{7}{2} \\ 0 & 1 & -\frac{1}{3} & 1 \\ 0 & 0 & \frac{5}{6} & \frac{5}{2} \end{array}\right]$$

Column 2 has now been transformed into

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

Observe that our transformation of column 2 did not change column 1.

To complete the Gauss–Jordan procedure, we must transform the third column of $A_6|\mathbf{b}_6$ into

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

We first use a Type 1 ERO to create a 1 in the third row and third column of $A_6|\mathbf{b}_6$. Then we use Type 2 EROs to put zeros in the rest of column 3. Steps 7–9 accomplish these goals.

Step 7 Multiply row 3 of $A_6|\mathbf{b}_6$ by $\frac{6}{5}$. The result of this Type 1 ERO is

$$A_7|\mathbf{b}_7 = \left[\begin{array}{ccc|c} 1 & 0 & \frac{5}{6} & \frac{7}{2} \\ 0 & 1 & -\frac{1}{3} & 1 \\ 0 & 0 & 3 & 3 \end{array}\right]$$

Step 8 Replace row 1 of $A_7|\mathbf{b}_7$ by $-\frac{5}{6}$(row 3 of $A_7|\mathbf{b}_7$) + row 1 of $A_7|\mathbf{b}_7$. The result of this Type 2 ERO is

$$A_8|\mathbf{b}_8 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{1}{3} & 1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Step 9 Replace row 2 of $A_8|\mathbf{b}_8$ by $\frac{1}{3}$(row 3 of $A_8|\mathbf{b}_8$) + row 2 of $A_8|\mathbf{b}_8$. The result of this Type 2 ERO is

$$A_9|\mathbf{b}_9 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$A_9|\mathbf{b}_9$ represents the system of equations

$$\begin{aligned} x_1 \qquad &= 1 \\ x_2 \quad &= 2 \\ x_3 &= 3 \end{aligned} \tag{9}$$

Thus, (9) has the unique solution $x_1 = 1, x_2 = 2, x_3 = 3$. Because (9) was obtained from (8) via EROs, the unique solution to (8) must also be $x_1 = 1, x_2 = 2, x_3 = 3$.

The reader might be wondering why we defined Type 3 EROs (interchanging of rows). To see why a Type 3 ERO might be useful, suppose you want to solve

$$\begin{aligned} 2x_2 + x_3 &= 6 \\ x_1 + x_2 - x_3 &= 2 \\ 2x_1 + x_2 + x_3 &= 4 \end{aligned} \tag{10}$$

To solve (10) by the Gauss–Jordan method, first form the augmented matrix

$$A|\mathbf{b} = \left[\begin{array}{ccc|c} 0 & 2 & 1 & 6 \\ 1 & 1 & -1 & 2 \\ 2 & 1 & 1 & 4 \end{array}\right]$$

The 0 in row 1 and column 1 means that a Type 1 ERO cannot be used to create a 1 in row 1 and column 1. If, however, we interchange rows 1 and 2 (a Type 3 ERO), we obtain

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 0 & 2 & 1 & 6 \\ 2 & 1 & 1 & 4 \end{array}\right] \tag{10$'$}$$

Now we may proceed as usual with the Gauss–Jordan method.

Special Cases: No Solution or an Infinite Number of Solutions

Some linear systems have no solution, and some have an infinite number of solutions. The following two examples illustrate how the Gauss–Jordan method can be used to recognize these cases.

EXAMPLE 6 **Linear System with No Solution**

Find all solutions to the following linear system:

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ 2x_1 + 4x_2 &= 4 \end{aligned} \tag{11}$$

Solution We apply the Gauss–Jordan method to the matrix

$$A|\mathbf{b} = \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 4 \end{array}\right]$$

We begin by replacing row 2 of $A|\mathbf{b}$ by -2(row 1 of $A|\mathbf{b}$) + row 2 of $A|\mathbf{b}$. The result of this Type 2 ERO is

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & -2 \end{array}\right] \tag{12}$$

We would now like to transform the second column of (12) into

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

but this is not possible. System (12) is equivalent to the following system of equations:

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ 0x_1 + 0x_2 &= -2 \end{aligned} \tag{12$'$}$$

Whatever values we give to x_1 and x_2, the second equation in (12′) can never be satisfied. Thus, (12′) has no solution. Because (12′) was obtained from (11) by use of EROs, (11) also has no solution.

Example 6 illustrates the following idea: *If you apply the Gauss–Jordan method to a linear system and obtain a row of the form* $[0 \quad 0 \quad \cdots \quad 0|c]$ $(c \neq 0)$, *then the original linear system has no solution.*

EXAMPLE 7 **Linear System with Infinite Number of Solutions**

Apply the Gauss–Jordan method to the following linear system:

$$\begin{aligned} x_1 + x_2 \phantom{{}+x_3} &= 1 \\ x_2 + x_3 &= 3 \\ x_1 + 2x_2 + x_3 &= 4 \end{aligned} \tag{13}$$

Solution The augmented matrix form of (13) is

$$A|\mathbf{b} = \left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 3 \\ 1 & 2 & 1 & 4 \end{array}\right]$$

We begin by replacing row 3 (because the row 2, column 1 value is already 0) of $A|\mathbf{b}$ by -1(row 1 of $A|\mathbf{b}$) + row 3 of $A|\mathbf{b}$. The result of this Type 2 ERO is

$$A_1|\mathbf{b}_1 = \left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 3 \\ 0 & 1 & 1 & 3 \end{array}\right] \tag{14}$$

Next we replace row 1 of $A_1|\mathbf{b}_1$ by -1(row 2 of $A_1|\mathbf{b}_1$) + row 1 of $A_1|\mathbf{b}_1$. The result of this Type 2 ERO is

$$A_2|\mathbf{b}_2 = \left[\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 1 & 1 & 3 \end{array}\right]$$

Now we replace row 3 of $A_2|\mathbf{b}_2$ by -1(row 2 of $A_2|\mathbf{b}_2$) + row 3 of $A_2|\mathbf{b}_2$. The result of this Type 2 ERO is

$$A_3|\mathbf{b}_3 = \left[\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We would now like to transform the third column of $A_3|\mathbf{b}_3$ into

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

but this is not possible. The linear system corresponding to $A_3|\mathbf{b}_3$ is

$$x_1 \qquad - x_3 = -2 \tag{14.1}$$

$$x_2 + x_3 = 3 \tag{14.2}$$

$$0x_1 + 0x_2 + 0x_3 = 0 \tag{14.3}$$

Suppose we assign an arbitrary value k to x_3. Then (14.1) will be satisfied if $x_1 - k = -2$, or $x_1 = k - 2$. Similarly, (14.2) will be satisfied if $x_2 + k = 3$, or $x_2 = 3 - k$. Of course, (14.3) will be satisfied for any values of x_1, x_2, and x_3. Thus, for any number k, $x_1 = k - 2$, $x_2 = 3 - k$, $x_3 = k$ is a solution to (14). Thus, (14) has an infinite number of solutions (one for each number k). Because (14) was obtained from (13) via EROs, (13) also has an infinite number of solutions. A more formal characterization of linear systems that have an infinite number of solutions will be given after the following summary of the Gauss–Jordan method.

Summary of the Gauss–Jordan Method

Step 1 To solve $A\mathbf{x} = \mathbf{b}$, write down the augmented matrix $A|\mathbf{b}$.

Step 2 At any stage, define a current row, current column, and current entry (the entry in the current row and column). Begin with row 1 as the current row, column 1 as the current column, and a_{11} as the current entry. (**a**) If a_{11} (the current entry) is nonzero, then use EROs to transform column 1 (the current column) to

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then obtain the new current row, column, and entry by moving down one row and one column to the right, and go to step 3. (**b**) If a_{11} (the current entry) equals 0, then do a Type 3 ERO involving the current row and any row that contains a nonzero number in the current column. Use EROs to transform column 1 to

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then obtain the new current row, column, and entry by moving down one row and one column to the right. Go to step 3. (**c**) If there are no nonzero numbers in the first column, then obtain a new current column and entry by moving one column to the right. Then go to step 3.

Step 3 (**a**) If the new current entry is nonzero, then use EROs to transform it to 1 and the rest of the current column's entries to 0. When finished, obtain the new current row, column, and entry. If this is impossible, then stop. Otherwise, repeat step 3. (**b**) If the current entry is 0, then do a Type 3 ERO with the current row and any row that contains a nonzero number in the current column. Then use EROs to transform that current entry to 1 and the rest of the current column's entries to 0. When finished, obtain the new current row, column, and entry. If this is impossible, then stop. Otherwise, repeat step 3. (**c**) If the current column has no nonzero numbers below the current row, then obtain the new current column and entry, and repeat step 3. If it is impossible, then stop.

This procedure may require "passing over" one or more columns without transforming them (see Problem 8).

Step 4 Write down the system of equations $A'\mathbf{x} = \mathbf{b}'$ that corresponds to the matrix $A'|\mathbf{b}'$ obtained when step 3 is completed. Then $A'\mathbf{x} = \mathbf{b}'$ will have the same set of solutions as $A\mathbf{x} = \mathbf{b}$.

Basic Variables and Solutions to Linear Equation Systems

To describe the set of solutions to $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$), we need to define the concepts of basic and nonbasic variables.

DEFINITION ■ After the Gauss–Jordan method has been applied to any linear system, a variable that appears with a coefficient of 1 in a single equation and a coefficient of 0 in all other equations is called a **basic variable** (BV). ■

Any variable that is not a basic variable is called a **nonbasic variable** (NBV). ■

Let *BV* be the set of basic variables for $A'\mathbf{x} = \mathbf{b}'$ and NBV be the set of nonbasic variables for $A'\mathbf{x} = \mathbf{b}'$. The character of the solutions to $A'\mathbf{x} = \mathbf{b}'$ depends on which of the following cases occurs.

Case 1 $A'\mathbf{x} = \mathbf{b}'$ has at least one row of form $[0 \quad 0 \quad \cdots \quad 0|c]$ $(c \neq 0)$. Then $A\mathbf{x} = \mathbf{b}$ has no solution (recall Example 6). As an example of Case 1, suppose that when the Gauss–Jordan method is applied to the system $A\mathbf{x} = \mathbf{b}$, the following matrix is obtained:

$$
A'|\mathbf{b}' = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{array}\right]
$$

In this case, $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$) has no solution.

Case 2 Suppose that Case 1 does not apply and NBV, the set of nonbasic variables, is empty. Then $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$) will have a unique solution. To illustrate this, we recall that in solving

$$
\begin{aligned}
2x_1 + 2x_2 + x_3 &= 9 \\
2x_1 - x_2 + 2x_3 &= 6 \\
x_1 - x_2 + 2x_3 &= 5
\end{aligned}
$$

the Gauss–Jordan method yielded

$$
A'|\mathbf{b}' = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]
$$

In this case, BV $= \{x_1, x_2, x_3\}$ and NBV is empty. Then the unique solution to $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$) is $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

Case 3 Suppose that Case 1 does not apply and NBV is nonempty. Then $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$) will have an infinite number of solutions. To obtain these, first assign each nonbasic variable an arbitrary value. Then solve for the value of each basic variable in terms of the nonbasic variables. For example, suppose

$$
A'|\mathbf{b}' = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 1 & 3 \\ 0 & 1 & 0 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{15}
$$

Because Case 1 does not apply, and BV $= \{x_1, x_2, x_3\}$ and NBV $= \{x_4, x_5\}$, we have an example of Case 3: $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$) will have an infinite number of solutions. To see what these solutions look like, write down $A'\mathbf{x} = \mathbf{b}'$:

$$
x_1 \qquad\qquad + x_4 + x_5 = 3 \tag{15.1}
$$

$$
x_2 \qquad + 2x_4 \qquad = 2 \tag{15.2}
$$

$$
x_3 \qquad + x_5 = 1 \tag{15.3}
$$

$$
0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = 0 \tag{15.4}
$$

Now assign the nonbasic variables (x_4 and x_5) arbitrary values c and k, with $x_4 = c$ and $x_5 = k$. From (15.1), we find that $x_1 = 3 - c - k$. From (15.2), we find that $x_2 = 2 - 2c$. From (15.3), we find that $x_3 = 1 - k$. Because (15.4) holds for all values of the variables, $x_1 = 3 - c - k$, $x_2 = 2 - 2c$, $x_3 = 1 - k$, $x_4 = c$, and $x_5 = k$ will, for any values of c and k, be a solution to $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$).

Our discussion of the Gauss–Jordan method is summarized in Figure 6. We have devoted so much time to the Gauss–Jordan method because, in our study of linear programming, examples of Case 3 (linear systems with an infinite number of solutions) will occur repeatedly. Because the end result of the Gauss–Jordan method must always be one of Cases 1–3, we have shown that any linear system will have no solution, a unique solution, or an infinite number of solutions.

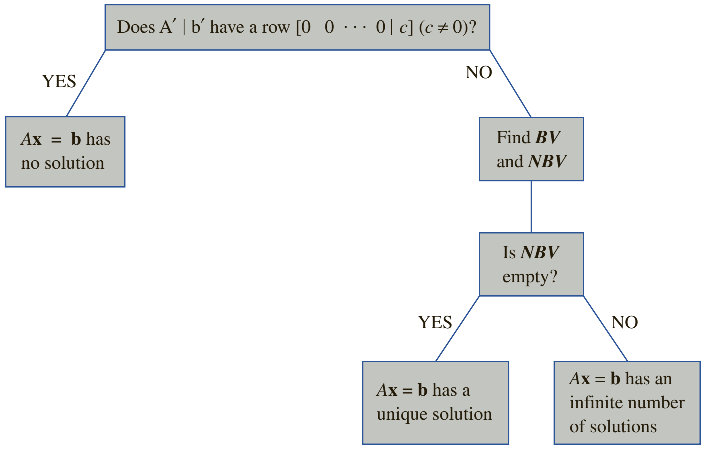


FIGURE 6
Description of Gauss–Jordan Method for Solving Linear Equations

PROBLEMS

Group A

Use the Gauss–Jordan method to determine whether each of the following linear systems has no solution, a unique solution, or an infinite number of solutions. Indicate the solutions (if any exist).

1
$$\begin{aligned} x_1 + x_2 + x_4 &= 3 \\ x_2 + x_3 &= 4 \\ x_1 + 2x_2 + x_3 + x_4 &= 8 \end{aligned}$$

2
$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ x_1 + 2x_2 &= 6 \end{aligned}$$

3
$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + x_2 &= 3 \\ 3x_1 + 2x_2 &= 4 \end{aligned}$$

4
$$\begin{aligned} 2x_1 - x_2 + x_3 + x_4 &= 6 \\ x_1 + x_2 + x_3 &= 4 \end{aligned}$$

5
$$\begin{aligned} x_1 + x_4 &= 5 \\ x_2 + 2x_4 &= 5 \\ x_3 + 0.5x_4 &= 1 \\ 2x_3 + x_4 &= 3 \end{aligned}$$

6
$$\begin{aligned} 2x_2 + 2x_3 &= 4 \\ x_1 + 2x_2 + x_3 &= 4 \\ x_2 - x_3 &= 0 \end{aligned}$$

7
$$\begin{aligned} x_1 + x_2 &= 2 \\ -x_2 + 2x_3 &= 3 \\ x_2 + x_3 &= 3 \end{aligned}$$

8
$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ x_2 + 2x_3 + x_4 &= 2 \\ x_4 &= 3 \end{aligned}$$

Group B

9 Suppose that a linear system $A\mathbf{x} = \mathbf{b}$ has more variables than equations. Show that $A\mathbf{x} = \mathbf{b}$ cannot have a unique solution.

2.4 Linear Independence and Linear Dependence†

In this section, we discuss the concepts of a linearly independent set of vectors, a linearly dependent set of vectors, and the rank of a matrix. These concepts will be useful in our study of matrix inverses.

Before defining a linearly independent set of vectors, we need to define a linear combination of a set of vectors. Let $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a set of row vectors all of which have the same dimension.

†This section covers topics that may be omitted with no loss of continuity.

DEFINITION ■ A **linear combination** of the vectors in V is any vector of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \cdots + c_k\mathbf{v}_k$, where $c_1, c_2, \ldots, c_k$ are arbitrary scalars. ■

For example, if $V = \{[1 \quad 2], [2 \quad 1]\}$, then

$$\begin{aligned} 2\mathbf{v}_1 - \mathbf{v}_2 &= 2([1 \quad 2]) - [2 \quad 1] = [0 \quad 3] \\ \mathbf{v}_1 + 3\mathbf{v}_2 &= [1 \quad 2] + 3([2 \quad 1]) = [7 \quad 5] \\ 0\mathbf{v}_1 + 3\mathbf{v}_2 &= [0 \quad 0] + 3([2 \quad 1]) = [6 \quad 3] \end{aligned}$$

are linear combinations of vectors in V. The foregoing definition may also be applied to a set of column vectors.

Suppose we are given a set $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of m-dimensional row vectors. Let $\mathbf{0} = [0 \quad 0 \quad \cdots \quad 0]$ be the m-dimensional $\mathbf{0}$ vector. To determine whether V is a linearly independent set of vectors, we try to find a linear combination of the vectors in V that adds up to $\mathbf{0}$. Clearly, $0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k$ is a linear combination of vectors in V that adds up to $\mathbf{0}$. We call the linear combination of vectors in V for which $c_1 = c_2 = \cdots = c_k = 0$ the *trivial* linear combination of vectors in V. We may now define linearly independent and linearly dependent sets of vectors.

DEFINITION ■ A set V of m-dimensional vectors is **linearly independent** if the only linear combination of vectors in V that equals $\mathbf{0}$ is the trivial linear combination. ■

A set V of m-dimensional vectors is **linearly dependent** if there is a nontrivial linear combination of the vectors in V that adds up to $\mathbf{0}$. ■

The following examples should clarify these definitions.

EXAMPLE 8 0 Vector Makes Set LD

Show that any set of vectors containing the $\mathbf{0}$ vector is a linearly dependent set.

Solution To illustrate, we show that if $V = \{[0 \quad 0], [1 \quad 0], [0 \quad 1]\}$, then V is linearly dependent, because if, say, $c_1 \neq 0$, then $c_1([0 \quad 0]) + 0([1 \quad 0]) + 0([0 \quad 1]) = [0 \quad 0]$. Thus, there is a nontrivial linear combination of vectors in V that adds up to $\mathbf{0}$.

EXAMPLE 9 LI Set of Vectors

Show that the set of vectors $V = \{[1 \quad 0], [0 \quad 1]\}$ is a linearly independent set of vectors.

Solution We try to find a nontrivial linear combination of the vectors in V that yields $\mathbf{0}$. This requires that we find scalars c_1 and c_2 (at least one of which is nonzero) satisfying $c_1([1 \quad 0]) + c_2([0 \quad 1]) = [0 \quad 0]$. Thus, c_1 and c_2 must satisfy $[c_1 \quad c_2] = [0 \quad 0]$. This implies $c_1 = c_2 = 0$. The only linear combination of vectors in V that yields $\mathbf{0}$ is the trivial linear combination. Therefore, V is a linearly independent set of vectors.

EXAMPLE 10 LD Set of Vectors

Show that $V = \{[1 \quad 2], [2 \quad 4]\}$ is a linearly dependent set of vectors.

Solution Because $2([1 \quad 2]) - 1([2 \quad 4]) = [0 \quad 0]$, there is a nontrivial linear combination with $c_1 = 2$ and $c_2 = -1$ that yields $\mathbf{0}$. Thus, V is a linearly dependent set of vectors.

Intuitively, what does it mean for a set of vectors to be linearly dependent? To understand the concept of linear dependence, observe that a set of vectors V is linearly dependent (as

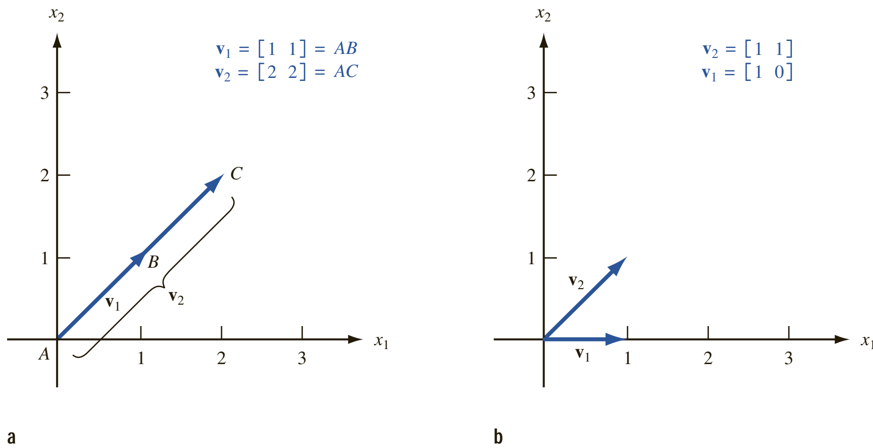


FIGURE 7
(a) Two Linearly Dependent Vectors (b) Two Linearly Independent Vectors

long as $\mathbf{0}$ is not in V) if and only if some vector in V can be written as a nontrivial linear combination of other vectors in V (see Problem 9 at the end of this section). For instance, in Example 10, $[2 \quad 4] = 2([1 \quad 2])$. Thus, if a set of vectors V is linearly dependent, the vectors in V are, in some way, not all "different" vectors. By "different" we mean that the direction specified by any vector in V cannot be expressed by adding together multiples of other vectors in V. For example, in two dimensions it can be shown that two vectors are linearly dependent if and only if they lie on the same line (see Figure 7).

The Rank of a Matrix

The Gauss–Jordan method can be used to determine whether a set of vectors is linearly independent or linearly dependent. Before describing how this is done, we define the concept of the rank of a matrix.

Let A be any $m \times n$ matrix, and denote the rows of A by $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$. Also define $R = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$.

DEFINITION ■ The **rank** of A is the number of vectors in the largest linearly independent subset of R. ■

The following three examples illustrate the concept of rank.

EXAMPLE 11 Matrix with 0 Rank

Show that rank $A = 0$ for the following matrix:

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

Solution For the set of vectors $R = \{[0 \quad 0], [0, \quad 0]\}$, it is impossible to choose a subset of R that is linearly independent (recall Example 8).

EXAMPLE 12 Matrix with Rank of 1

Show that rank $A = 1$ for the following matrix:

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

Solution Here $R = \{[1 \quad 1], [2 \quad 2]\}$. The set $\{[1 \quad 1]\}$ is a linearly independent subset of R, so rank A must be at least 1. If we try to find two linearly independent vectors in R, we fail because $2([1 \quad 1]) - [2 \quad 2] = [0 \quad 0]$. This means that rank A cannot be 2. Thus, rank A must equal 1.

EXAMPLE 13 **Matrix with Rank of 2**

Show that rank $A = 2$ for the following matrix:

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

Solution Here $R = \{[1 \quad 0], [0 \quad 1]\}$. From Example 9, we know that R is a linearly independent set of vectors. Thus, rank $A = 2$.

To find the rank of a given matrix A, simply apply the Gauss–Jordan method to the matrix A. Let the final result be the matrix $\bar{A}$. It can be shown that performing a sequence of EROs on a matrix does not change the rank of the matrix. This implies that rank $A =$ rank $\bar{A}$. It is also apparent that the rank of $\bar{A}$ will be the number of nonzero rows in $\bar{A}$. Combining these facts, we find that rank $A =$ rank $\bar{A} =$ number of nonzero rows in $\bar{A}$.

EXAMPLE 14 **Using Gauss–Jordan Method to Find Rank of Matrix**

Find

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

Solution The Gauss–Jordan method yields the following sequence of matrices:

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

$$= \bar{A}$$

Thus, rank $A =$ rank $\bar{A} = 3$.

How to Tell Whether a Set of Vectors Is Linearly Independent

We now describe a method for determining whether a set of vectors $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is linearly independent.

Form the matrix A whose ith row is $\mathbf{v}_i$. A will have m rows. If rank $A = m$, then V is a linearly independent set of vectors, whereas if rank $A < m$, then V is a linearly dependent set of vectors.

EXAMPLE 15 **A Linearly Dependent Set of Vectors**

Determine whether $V = \{[1 \quad 0 \quad 0], [0 \quad 1 \quad 0], [1 \quad 1 \quad 0]\}$ is a linearly independent set of vectors.

Solution The Gauss–Jordan method yields the following sequence of matrices:

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

Thus, rank A = rank $\bar{A} = 2 < 3$. This shows that V is a linearly dependent set of vectors. In fact, the EROs used to transform A to $\bar{A}$ can be used to show that $[1 \quad 1 \quad 0] = [1 \quad 0 \quad 0] + [0 \quad 1 \quad 0]$. This equation also shows that V is a linearly dependent set of vectors.

PROBLEMS

Group A

Determine if each of the following sets of vectors is linearly independent or linearly dependent.

1 $V = \{[1 \quad 0 \quad 1], [1 \quad 2 \quad 1], [2 \quad 2 \quad 2]\}$

2 $V = \{[2 \quad 1 \quad 0], [1 \quad 2 \quad 0], [3 \quad 3 \quad 1]\}$

3 $V = \{[2 \quad 1], [1 \quad 2]\}$

4 $V = \{[2 \quad 0], [3 \quad 0]\}$

5 $V = \left\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}\right\}$

6 $V = \left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$

Group B

7 Show that the linear system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ can be written as a linear combination of the columns of A.

8 Suppose there is a collection of three or more two-dimensional vectors. Provide an argument showing that the collection must be linearly dependent.

9 Show that a set of vectors V (not containing the $\mathbf{0}$ vector) is linearly dependent if and only if there exists some vector in V that can be written as a nontrivial linear combination of other vectors in V.

2.5 The Inverse of a Matrix

To solve a single linear equation such as $4x = 3$, we simply multiply both sides of the equation by the multiplicative inverse of 4, which is 4^{-1}, or $\frac{1}{4}$. This yields $4^{-1}(4x) = (4^{-1})3$, or $x = \frac{3}{4}$. (Of course, this method fails to work for the equation $0x = 3$, because zero has no multiplicative inverse.) In this section, we develop a generalization of this technique that can be used to solve "square" (number of equations = number of unknowns) linear systems. We begin with some preliminary definitions.

DEFINITION ■ A **square matrix** is any matrix that has an equal number of rows and columns. ■

The **diagonal elements** of a square matrix are those elements a_{ij} such that $i = j$. ■

A square matrix for which all diagonal elements are equal to 1 and all nondiagonal elements are equal to 0 is called an **identity matrix.** ■

The $m \times m$ identity matrix will be written as I_m. Thus,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \cdots$$

If the multiplications $I_m A$ and AI_m are defined, it is easy to show that $I_m A = AI_m = A$. Thus, just as the number 1 serves as the unit element for multiplication of real numbers, I_m serves as the unit element for multiplication of matrices.

Recall that $\frac{1}{4}$ is the multiplicative inverse of 4. This is because $4(\frac{1}{4}) = (\frac{1}{4})4 = 1$. This motivates the following definition of the inverse of a matrix.

DEFINITION ■ For a given $m \times m$ matrix A, the $m \times m$ matrix B is the **inverse** of A if

$$BA = AB = I_m \tag{16}$$

(It can be shown that if $BA = I_m$ or $AB = I_m$, then the other quantity will also equal I_m.) ■

Some square matrices do not have inverses. If there does exist an $m \times m$ matrix B that satisfies Equation (16), then we write $B = A^{-1}$. For example, if

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

the reader can verify that

$$\begin{bmatrix} 2 & 0 & -1 \\ 3 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ -5 & 1 & -7 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 & 1 \\ -5 & 1 & -7 \\ 1 & 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & 0 & -1 \\ 3 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus,

$$A^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ -5 & 1 & -7 \\ 1 & 0 & 2 \end{bmatrix}$$

To see why we are interested in the concept of a matrix inverse, suppose we want to solve a linear system $A\mathbf{x} = \mathbf{b}$ that has m equations and m unknowns. Suppose that A^{-1} exists. Multiplying both sides of $A\mathbf{x} = \mathbf{b}$ by A^{-1}, we see that any solution of $A\mathbf{x} = \mathbf{b}$ must also satisfy $A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b}$. Using the associative law and the definition of a matrix inverse, we obtain

$$(A^{-1}A)\mathbf{x} = A^{-1}\mathbf{b}$$

or $$I_m\mathbf{x} = A^{-1}\mathbf{b}$$

or $$\mathbf{x} = A^{-1}\mathbf{b}$$

This shows that knowing A^{-1} enables us to find the unique solution to a square linear system. This is the analog of solving $4x = 3$ by multiplying both sides of the equation by 4^{-1}.

The Gauss–Jordan method may be used to find A^{-1} (or to show that A^{-1} does not exist). To illustrate how we can use the Gauss–Jordan method to invert a matrix, suppose we want to find A^{-1} for

$$A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

This requires that we find a matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = A^{-1}$$

that satisfies

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{17}$$

From Equation (17), we obtain the following pair of simultaneous equations that must be satisfied by a, b, c, and d:

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \qquad \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Thus, to find

$$\begin{bmatrix} a \\ c \end{bmatrix}$$

(the first column of A^{-1}), we can apply the Gauss–Jordan method to the augmented matrix

$$\left[\begin{array}{cc|c} 2 & 5 & 1 \\ 1 & 3 & 0 \end{array}\right]$$

Once EROs have transformed

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

to I_2,

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

will have been transformed into the first column of A^{-1}. To determine

$$\begin{bmatrix} b \\ d \end{bmatrix}$$

(the second column of A^{-1}), we apply EROs to the augmented matrix

$$\left[\begin{array}{cc|c} 2 & 5 & 0 \\ 1 & 3 & 1 \end{array}\right]$$

When

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

has been transformed into I_2,

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

will have been transformed into the second column of A^{-1}. Thus, to find each column of A^{-1}, we must perform a sequence of EROs that transform

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

into I_2. This suggests that we can find A^{-1} by applying EROs to the 2×4 matrix

$$A|I_2 = \left[\begin{array}{cc|cc} 2 & 5 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right]$$

When

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

has been transformed to I_2,

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

will have been transformed into the first column of A^{-1}, and

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

will have been transformed into the second column of A^{-1}. Thus, *as A is transformed into I_2, I_2 is transformed into A^{-1}*. The computations to determine A^{-1} follow.

Step 1 Multiply row 1 of $A|I_2$ by $\frac{1}{2}$. This yields

$$A'|I_2' = \left[\begin{array}{cc|cc} 1 & \frac{5}{2} & \frac{1}{2} & 0 \\ 1 & 3 & 0 & 1 \end{array}\right]$$

Step 2 Replace row 2 of $A'|I_2'$ by -1(row 1 of $A'|I_2'$) + row 2 of $A'|I_2'$. This yields

$$A''|I_2'' = \left[\begin{array}{cc|cc} 1 & \frac{5}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 1 \end{array}\right]$$

Step 3 Multiply row 2 of $A''|I_2''$ by 2. This yields

$$A'''|I_2''' = \left[\begin{array}{cc|cc} 1 & \frac{5}{2} & \frac{1}{2} & 0 \\ 0 & 1 & -1 & 2 \end{array}\right]$$

Step 4 Replace row 1 of $A'''|I_2'''$ by $-\frac{5}{2}$(row 2 of $A'''|I_2'''$) + row 1 of $A'''|I_2'''$. This yields

$$\left[\begin{array}{cc|cc} 1 & 0 & 3 & -5 \\ 0 & 1 & -1 & 2 \end{array}\right]$$

Because A has been transformed into I_2, I_2 will have been transformed into A^{-1}. Hence,

$$A^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$$

The reader should verify that $AA^{-1} = A^{-1}A = I_2$.

A Matrix May Not Have an Inverse

Some matrices do not have inverses. To illustrate, let

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \tag{18}$$

To find A^{-1} we must solve the following pair of simultaneous equations:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} e \\ g \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{18.1}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} f \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{18.2}$$

When we try to solve (18.1) by the Gauss–Jordan method, we find that

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 2 & 4 & 0 \end{array}\right]$$

is transformed into

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 0 & -2 \end{array}\right]$$

This indicates that (18.1) has no solution, and A^{-1} cannot exist.

Observe that (18.1) fails to have a solution, because the Gauss–Jordan method transforms A into a matrix with a row of zeros on the bottom. This can only happen if rank $A < 2$. If $m \times m$ matrix A has rank $A < m$, then A^{-1} will not exist.

The Gauss–Jordan Method for Inverting an $m \times m$ Matrix A

Step 1 Write down the $m \times 2m$ matrix $A|I_m$.

Step 1 Use EROs to transform $A|I_m$ into $I_m|B$. This will be possible only if rank $A = m$. In this case, $B = A^{-1}$. If rank $A < m$, then A has no inverse.

Using Matrix Inverses to Solve Linear Systems

As previously stated, matrix inverses can be used to solve a linear system $A\mathbf{x} = \mathbf{b}$ in which the number of variables and equations are equal. Simply multiply both sides of $A\mathbf{x} = \mathbf{b}$ by A^{-1} to obtain the solution $\mathbf{x} = A^{-1}\mathbf{b}$. For example, to solve

$$\begin{aligned} 2x_1 + 5x_2 &= 7 \\ x_1 + 3x_2 &= 4 \end{aligned} \tag{19}$$

write the matrix representation of (19):

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix} \tag{20}$$

Let

$$A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

We found in the previous illustration that

$$A^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$$

	A	B	C	D	E	F	G	H
1		**Inverting**						
2		**a**						
3		**Matrix**			2	0	-1	
4				A	3	1	2	
5					-1	0	1	
6								
7					1	0	1	
8				A^{-1}	-5	1	-7	
9					1	0	2	

FIGURE 8

Multiplying both sides of (20) by A^{-1}, we obtain

$$\begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 7 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus, $x_1 = 1$, $x_2 = 1$ is the unique solution to system (19).

Inverting Matrices with Excel

Minverse.xls

The Excel =MINVERSE command makes it easy to invert a matrix. See Figure 8 and file Minverse.xls. Suppose we want to invert the matrix

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

Simply enter the matrix in E3:G5 and select the range (we chose E7:G9) where you want A^{-1} to be computed. In the upper left-hand corner of the range E7:G9 (cell E7), we enter the formula

= MINVERSE(E3:G5)

and select **Control Shift Enter.** This enters an array function that computes A^{-1} in the range E7:G9. You cannot edit part of an array function, so if you want to delete A^{-1}, you must delete the entire range where A^{-1} is present.

PROBLEMS

Group A

Find A^{-1} (if it exists) for the following matrices:

1 $\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$

2 $\begin{bmatrix} 1 & 0 & 1 \\ 4 & 1 & -2 \\ 3 & 1 & -1 \end{bmatrix}$

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

4 $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 0 \\ 2 & 4 & 1 \end{bmatrix}$

5 Use the answer to Problem 1 to solve the following linear system:

$$\begin{aligned} x_1 + 3x_2 &= 4 \\ 2x_1 + 5x_2 &= 7 \end{aligned}$$

6 Use the answer to Problem 2 to solve the following linear system:

$$\begin{aligned} x_1 + x - 2x_3 &= 4 \\ 4x_1 + x_2 - 2x_3 &= 0 \\ 3x_1 + x_2 - x_3 &= 2 \end{aligned}$$

Group B

7 Show that a square matrix has an inverse if and only if its rows form a linearly independent set of vectors.

8 Consider a square matrix B whose inverse is given by B^{-1}.

a In terms of B^{-1}, what is the inverse of the matrix $100B$?

b Let B' be the matrix obtained from B by doubling every entry in row 1 of B. Explain how we could obtain the inverse of B' from B^{-1}.

c Let B' be the matrix obtained from B by doubling every entry in column 1 of B. Explain how we could obtain the inverse of B' from B^{-1}.

9 Suppose that A and B both have inverses. Find the inverse of the matrix AB.

10 Suppose A has an inverse. Show that $(A^T)^{-1} = (A^{-1})^T$. (*Hint:* Use the fact that $AA^{-1} = I$, and take the transpose of both sides.)

11 A square matrix A is *orthogonal* if $AA^T = I$. What properties must be possessed by the columns of an orthogonal matrix?

2.6 Determinants

Associated with any square matrix A is a number called the *determinant* of A (often abbreviated as det A or $|A|$). Knowing how to compute the determinant of a square matrix will be useful in our study of nonlinear programming.

For a 1×1 matrix $A = [a_{11}]$,

$$\det A = a_{11} \tag{21}$$

For a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{22}$$

$$\det A = a_{11}a_{22} - a_{21}a_{12}$$

For example,

$$\det \begin{bmatrix} 2 & 4 \\ 3 & 5 \end{bmatrix} = 2(5) - 3(4) = -2$$

Before we learn how to compute det A for larger square matrices, we need to define the concept of the *minor* of a matrix.

DEFINITION ■ If A is an $m \times m$ matrix, then for any values of i and j, the ijth **minor** of A (written A_{ij}) is the $(m - 1) \times (m - 1)$ submatrix of A obtained by deleting row i and column j of A. ■

For example,

$$\text{if} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \text{then} \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} \quad \text{and} \quad A_{32} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

Let A be any $m \times m$ matrix. We may write A as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

To compute det A, pick any value of i ($i = 1, 2, \ldots, m$) and compute det A:

$$\det A = (-1)^{i+1}a_{i1}(\det A_{i1}) + (-1)^{i+2}a_{i2}(\det A_{i2}) + \cdots + (-1)^{i+m}a_{im}(\det A_{im}) \tag{23}$$

Formula (23) is called the expansion of det A by the cofactors of row i. The virtue of (23) is that it reduces the computation of det A for an $m \times m$ matrix to computations involving only $(m - 1) \times (m - 1)$ matrices. Apply (23) until det A can be expressed in terms of 2×2 matrices. Then use Equation (22) to find the determinants of the relevant 2×2 matrices.

To illustrate the use of (23), we find det A for

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

We expand det A by using row 1 cofactors. Notice that $a_{11} = 1$, $a_{12} = 2$, and $a_{13} = 3$. Also

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}$$

so by (22), det $A_{11} = 5(9) - 8(6) = -3$;

$$A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}$$

so by (22), det $A_{12} = 4(9) - 7(6) = -6$; and

$$A_{13} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

so by (22), det $A_{13} = 4(8) - 7(5) = -3$. Then by (23),

$$\begin{aligned} \det A &= (-1)^{1+1}a_{11}(\det A_{11}) + (-1)^{1+2}a_{12}(\det A_{12}) + (-1)^{1+3}a_{13}(\det A_{13}) \\ &= (1)(1)(-3) + (-1)(2)(-6) + (1)(3)(-3) = -3 + 12 - 9 = 0 \end{aligned}$$

The interested reader may verify that expansion of det A by either row 2 or row 3 cofactors also yields det $A = 0$.

We close our discussion of determinants by noting that they can be used to invert square matrices and to solve linear equation systems. Because we already have learned to use the Gauss–Jordan method to invert matrices and to solve linear equation systems, we will not discuss these uses of determinants.

PROBLEMS

Group A

1 Verify that $\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = 0$ by using expansions by row 2 and row 3 cofactors.

2 Find $\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$

3 A matrix is said to be upper triangular if for $i > j$, $a_{ij} = 0$. Show that the determinant of any upper triangular 3×3 matrix is equal to the product of the matrix's diagonal elements. (This result is true for any upper triangular matrix.)

Group B

4 **a** Show that for any 1×1 and 3×3 matrix, det $-A = -$det A.

b Show that for any 2×2 and 4×4 matrix, det $-A =$ det A.

c Generalize the results of parts (a) and (b).

SUMMARY

Matrices

A **matrix** is any rectangular array of numbers. For the matrix A, we let a_{ij} represent the element of A in row i and column j.

A matrix with only one row or one column may be thought of as a **vector**. Vectors appear in boldface type ($\mathbf{v}$). Given a row vector $\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n]$ and a column

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

of the same dimension, the **scalar product** of $\mathbf{u}$ and $\mathbf{v}$ (written $\mathbf{u} \cdot \mathbf{v}$) is the number $u_1v_1 + u_2v_2 + \cdots + u_nv_n$.

Given two matrices A and B, the **matrix product** of A and B (written AB) is defined if and only if the number of columns in $A =$ the number of rows in B. Suppose this is the case and A has m rows and B has n columns. Then the matrix product $C = AB$ of A and B is the $m \times n$ matrix C whose ijth element is determined as follows: The ijth element of $C =$ the scalar product of row i of A with column j of B.

Matrices and Linear Equations

The **linear equation system**

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad\qquad \vdots \quad &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

may be written as $A\mathbf{x} = \mathbf{b}$ or $A|\mathbf{b}$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

The Gauss–Jordan Method

Using **elementary row operations** (EROs), we may solve any linear equation system. From a matrix A, an ERO yields a new matrix A' via one of three procedures.

Type 1 ERO

Obtain A' by multiplying any row of A by a nonzero scalar.

Type 2 ERO

Multiply any row of A (say, row i) by a nonzero scalar c. For some $j \neq i$, let row j of $A' = c(\text{row } i \text{ of } A) + \text{row } j$ of A, and let the other rows of A' be the same as the rows of A.

Type 3 ERO

Interchange any two rows of A.

The Gauss–Jordan method uses EROs to solve linear equation systems, as shown in the following steps.

Step 1 To solve $A\mathbf{x} = \mathbf{b}$, write down the augmented matrix $A|\mathbf{b}$.

Step 2 Begin with row 1 as the current row, column 1 as the current column, and a_{11} as the current entry. (**a**) If a_{11} (the current entry) is nonzero, then use EROs to transform column 1 (the current column) to

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then obtain the new current row, column, and entry by moving down one row and one column to the right, and go to step 3. (**b**) If a_{11} (the current entry) equals 0, then do a Type 3 ERO switch with any row with a nonzero value in the same column. Use EROs to transform column 1 to

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and proceed to step 3 after moving into a new current row, column, and entry. (**c**) If there are no nonzero numbers in the first column, then proceed to a new current column and entry. Then go to step 3.

Step 3 (**a**) If the current entry is nonzero, use EROs to transform it to 1 and the rest of the current column's entries to 0. Obtain the new current row, column, and entry. If this is impossible, then stop. Otherwise, repeat step 3. (**b**) If the current entry is 0, then do a Type 3 ERO switch with any row with a nonzero value in the same column. Transform the column using EROs and move to the next current entry. If this is impossible, then stop. Otherwise, repeat step 3. (**c**) If the current column has no nonzero numbers below the current row, then obtain the new current column and entry, and repeat step 3. If it is impossible, then stop.

This procedure may require "passing over" one or more columns without transforming them.

Step 4 Write down the system of equations $A'\mathbf{x} = \mathbf{b}'$ that corresponds to the matrix $A'|\mathbf{b}'$ obtained when step 3 is completed. Then $A'\mathbf{x} = \mathbf{b}'$ will have the same set of solutions as $A\mathbf{x} = \mathbf{b}$.

To describe the set of solutions to $A'\mathbf{x} = \mathbf{b}'$ (and $A\mathbf{x} = \mathbf{b}$), we define the concepts of basic and nonbasic variables. After the Gauss–Jordan method has been applied to any linear system, a variable that appears with a coefficient of 1 in a single equation and a coefficient of 0 in all other equations is called a **basic variable.** Any variable that is not a basic variable is called a **nonbasic variable.**

Let BV be the set of basic variables for $A'\mathbf{x} = \mathbf{b}'$ and NBV be the set of nonbasic variables for $A'\mathbf{x} = \mathbf{b}'$.

Case 1 $A'\mathbf{x} = \mathbf{b}'$ contains at least one row of the form $[0 \quad 0 \quad \cdots \quad 0|c] (c \neq 0)$. In this case, $A\mathbf{x} = \mathbf{b}$ has no solution.

Case 2 If Case 1 does not apply and NBV, the set of nonbasic variables, is empty, then $A\mathbf{x} = \mathbf{b}$ will have a unique solution.

Case 3 If Case 1 does not hold and NBV is nonempty, then $A\mathbf{x} = \mathbf{b}$ will have an infinite number of solutions.

Linear Independence, Linear Dependence, and the Rank of a Matrix

A set V of m-dimensional vectors is **linearly independent** if the only linear combination of vectors in V that equals $\mathbf{0}$ is the trivial linear combination. A set V of m-dimensional vectors is **linearly dependent** if there is a nontrivial linear combination of the vectors in V that adds to $\mathbf{0}$.

Let A be any $m \times n$ matrix, and denote the rows of A by $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$. Also define $R = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$. The **rank** of A is the number of vectors in the largest linearly independent subset of R. To find the rank of a given matrix A, apply the Gauss–Jordan method to the matrix A. Let the final result be the matrix $\bar{A}$. Then rank A = rank $\bar{A}$ = number of nonzero rows in $\bar{A}$.

To determine if a set of vectors $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is linearly dependent, form the matrix A whose ith row is $\mathbf{v}_i$. A will have m rows. If rank $A = m$, then V is a linearly independent set of vectors; if rank $A < m$, then V is a linearly dependent set of vectors.

Inverse of a Matrix

For a given square ($m \times m$) matrix A, if $AB = BA = I_m$, then B is the **inverse** of A (written $B = A^{-1}$). The Gauss–Jordan method for inverting an $m \times m$ matrix A to get A^{-1} is as follows:

Step 1 Write down the $m \times 2m$ matrix $A|I_m$.

Step 2 Use EROs to transform $A|I_m$ into $I_m|B$. This will only be possible if rank $A = m$. In this case, $B = A^{-1}$. If rank $A < m$, then A has no inverse.

Determinants

Associated with any square ($m \times m$) matrix A is a number called the **determinant** of A (written det A or $|A|$). For a 1×1 matrix, det $A = a_{11}$. For a 2×2 matrix, det $A = a_{11}a_{22} - a_{21}a_{12}$. For a general $m \times m$ matrix, we can find det A by repeated application of the following formula (valid for $i = 1, 2, \ldots, m$):

$$\det A = (-1)^{i+1}a_{i1}(\det A_{i1}) + (-1)^{i+2}a_{i2}(\det A_{i2}) + \cdots + (-1)^{i+m}a_{im}(\det A_{im})$$

Here A_{ij} is the ijth **minor** of A, which is the $(m-1) \times (m-1)$ matrix obtained from A after deleting the ith row and jth column of A.

REVIEW PROBLEMS

Group A

1 Find all solutions to the following linear system:

$$\begin{aligned} x_1 + \quad x_2 \quad\quad &= 2 \\ x_2 + x_3 &= 3 \\ x_1 + 2x_2 + x_3 &= 5 \end{aligned}$$

2 Find the inverse of the matrix $\begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix}$.

3 Each year, 20% of all untenured State University faculty become tenured, 5% quit, and 75% remain untenured. Each year, 90% of all tenured S.U. faculty remain tenured and 10% quit. Let U_t be the number of untenured S.U. faculty at the beginning of year t, and T_t the tenured number.

Use matrix multiplication to relate the vector $\begin{bmatrix} U_{t+1} \\ T_{t+1} \end{bmatrix}$ to the vector $\begin{bmatrix} U_t \\ T_t \end{bmatrix}$.

4 Use the Gauss–Jordan method to determine all solutions to the following linear system:

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ x_1 + \quad x_2 &= 1 \\ x_1 + 2x_2 &= 2 \end{aligned}$$

5 Find the inverse of the matrix $\begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix}$.

6 The grades of two students during their last semester at S.U. are shown in Table 2.

Courses 1 and 2 are four-credit courses, and courses 3 and 4 are three-credit courses. Let GPA_i be the semester grade point average for student i. Use matrix multiplication to express the vector $\begin{bmatrix} \text{GPA}_1 \\ \text{GPA}_2 \end{bmatrix}$ in terms of the information given in the problem.

7 Use the Gauss–Jordan method to find all solutions to the following linear system:

$$\begin{aligned} 2x_1 + x_2 &= 3 \\ 3x_1 + x_2 &= 4 \\ x_1 - x_2 &= 0 \end{aligned}$$

8 Find the inverse of the matrix $\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$.

9 Let C_t = number of children in Indiana at the beginning of year t, and A_t = number of adults in Indiana at the beginning of year t. During any given year, 5% of all children become adults, and 1% of all children die. Also, during any given year, 3% of all adults die. Use matrix multiplication to express the vector $\begin{bmatrix} C_{t+1} \\ A_{t+1} \end{bmatrix}$ in terms of $\begin{bmatrix} C_t \\ A_t \end{bmatrix}$.

TABLE 2

Student	Course 1	2	3	4
1	3.6	3.8	2.6	3.4
2	2.7	3.1	2.9	3.6

10 Use the Gauss–Jordan method to find all solutions to the following linear equation system:

$$\begin{aligned} x_1 \quad\quad - x_3 &= 4 \\ x_2 + x_3 &= 2 \\ x_1 + x_2 \quad\quad &= 5 \end{aligned}$$

11 Use the Gauss–Jordan method to find the inverse of the matrix $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

12 During any given year, 10% of all rural residents move to the city, and 20% of all city residents move to a rural area (all other people stay put!). Let R_t be the number of rural residents at the beginning of year t, and C_t be the number of city residents at the beginning of year t. Use matrix multiplication to relate the vector $\begin{bmatrix} R_{t+1} \\ C_{t+1} \end{bmatrix}$ to the vector $\begin{bmatrix} R_t \\ C_t \end{bmatrix}$.

13 Determine whether the set $V = \{[1 \quad 2 \quad 1], [2 \quad 0 \quad 0]\}$ is a linearly independent set of vectors.

14 Determine whether the set $V = \{[1 \quad 0 \quad 0], [0 \quad 1 \quad 0], [-1 \quad -1 \quad 0]\}$ is a linearly independent set of vectors.

15 Let $A = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{bmatrix}$.

a For what values of a, b, c, and d will A^{-1} exist?

b If A^{-1} exists, then find it.

16 Show that the following linear system has an infinite number of solutions:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix}$$

17 Before paying employee bonuses and state and federal taxes, a company earns profits of $60,000. The company pays employees a bonus equal to 5% of after-tax profits. State tax is 5% of profits (after bonuses are paid). Finally, federal tax is 40% of profits (after bonuses and state tax are paid). Determine a linear equation system to find the amounts paid in bonuses, state tax, and federal tax.

18 Find the determinant of the matrix $A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

19 Show that any 2×2 matrix A that does not have an inverse will have $\det A = 0$.

Group B

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

a Show that if rank $A = m$, then $A\mathbf{x} = \mathbf{0}$ has a unique solution. What is the unique solution?

b Show that if rank $A < m$, then $A\mathbf{x} = \mathbf{0}$ has an infinite number of solutions.

21 Consider the following linear system:

$$[x_1 \quad x_2 \quad \cdots \quad x_n] = [x_1 \quad x_2 \quad \cdots \quad x_n]P$$

where

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

If the sum of each row of the P matrix equals 1, then use Problem 20 to show that this linear system has an infinite number of solutions.

22[†] The national economy of Seriland manufactures three products: steel, cars, and machines. (1) To produce \$1 of steel requires 30¢ of steel, 15¢ of cars, and 40¢ of machines. (2) To produce \$1 of cars requires 45¢ of steel, 20¢ of cars, and 10¢ of machines. (3) To produce \$1 of machines requires 40¢ of steel, 10¢ of cars, and 45¢ of machines. During the coming year, Seriland wants to consume d_s dollars of steel, d_c dollars of cars, and d_m dollars of machinery.

For the coming year, let

s = dollar value of steel produced
c = dollar value of cars produced
m = dollar value of machines produced

Define A to be the 3×3 matrix whose ijth element is the dollar value of product i required to produce \$1 of product j (steel = product 1, cars = product 2, machinery = product 3).

a Determine A.

b Show that

$$\begin{bmatrix} s \\ c \\ m \end{bmatrix} = A\begin{bmatrix} s \\ c \\ m \end{bmatrix} + \begin{bmatrix} d_s \\ d_c \\ d_m \end{bmatrix} \tag{24}$$

(*Hint:* Observe that the value of next year's steel production = (next year's consumer steel demand) + (steel needed to make next year's steel) + (steel needed to make next year's cars) + (steel needed to make next year's machines). This should give you the general idea.)

c Show that Equation (24) may be rewritten as

$$(I - A)\begin{bmatrix} s \\ c \\ m \end{bmatrix} = \begin{bmatrix} d_s \\ d_c \\ d_m \end{bmatrix}$$

d Given values for d_s, d_c, and d_m, describe how you can use $(I - A)^{-1}$ to determine if Seriland can meet next year's consumer demand.

e Suppose next year's demand for steel increases by \$1. This will increase the value of the steel, cars, and machines that must be produced next year. In terms of $(I - A)^{-1}$, determine the change in next year's production requirements.

REFERENCES

The following references contain more advanced discussions of linear algebra. To understand the theory of linear and nonlinear programming, master at least one of these books:

Dantzig, G. *Linear Programming and Extensions.* Princeton, N.J.: Princeton University Press, 1963.
Hadley, G. *Linear Algebra.* Reading, Mass.: Addison-Wesley, 1961.
Strang, G. *Linear Algebra and Its Applications,* 3d ed. Orlando, Fla.: Academic Press, 1988.
Leontief, W. *Input–Output Economics.* New York: Oxford University Press, 1966.
Teichroew, D. *An Introduction to Management Science: Deterministic Models.* New York: Wiley, 1964. A more extensive discussion of linear algebra than this chapter gives (at a comparable level of difficulty).

[†]Based on Leontief (1966). See references at end of chapter.

Properties of matrix operations

The operations are as follows:

- Addition: if A and B are matrices of the same size $m \times n$, then $A + B$, their sum, is a matrix of size $m \times n$.
- Multiplication by scalars: if A is a matrix of size $m \times n$ and c is a scalar, then cA is a matrix of size $m \times n$.
- Matrix multiplication: if A is a matrix of size $m \times n$ and B is a matrix of size $n \times p$, then the product AB is a matrix of size $m \times p$.
- Vectors: a vector of length n can be treated as a matrix of size $n \times 1$, and the operations of vector addition, multiplication by scalars, and multiplying a matrix by a vector agree with the corresponding matrix operations.
- Transpose: if A is a matrix of size $m \times n$, then its transpose A^T is a matrix of size $n \times m$.
- Identity matrix: I_n is the $n \times n$ identity matrix; its diagonal elements are equal to 1 and its offdiagonal elements are equal to 0.
- Zero matrix: we denote by 0 the matrix of all zeroes (of relevant size).
- Inverse: if A is a **square** matrix, then its inverse A^{-1} is a matrix of the same size. Not every square matrix has an inverse! (The matrices that have inverses are called **invertible**.)

The properties of these operations are (assuming that r, s are scalars and the sizes of the matrices A, B, C are chosen so that each operation is well defined):

$$A + B = B + A, \tag{1}$$
$$(A + B) + C = A + (B + C), \tag{2}$$
$$A + 0 = A, \tag{3}$$
$$r(A + B) = rA + rB, \tag{4}$$
$$(r + s)A = rA + sA, \tag{5}$$
$$r(sA) = (rs)A; \tag{6}$$
$$A(BC) = (AB)C, \tag{7}$$
$$A(B + C) = AB + AC, \tag{8}$$
$$(B + C)A = BA + CA, \tag{9}$$

$$r(AB) = (rA)B = A(rB), \tag{10}$$
$$I_m A = A = A I_n; \tag{11}$$
$$(A^T)^T = A, \tag{12}$$
$$(A+B)^T = A^T + B^T, \tag{13}$$
$$(rA)^T = rA^T, \tag{14}$$
$$(AB)^T = B^T A^T, \tag{15}$$
$$(I_n)^T = I_n; \tag{16}$$
$$AA^{-1} = A^{-1}A = I_n, \tag{17}$$
$$(rA)^{-1} = r^{-1}A^{-1},\ r \neq 0, \tag{18}$$
$$(AB)^{-1} = B^{-1}A^{-1}, \tag{19}$$
$$(I_n)^{-1} = I_n, \tag{20}$$
$$(A^T)^{-1} = (A^{-1})^T, \tag{21}$$
$$(A^{-1})^{-1} = A. \tag{22}$$

We see that in many cases, we can treat addition and multiplication of matrices as addition and multiplication of numbers. However, here are some differences between operations with matrices and operations with numbers:

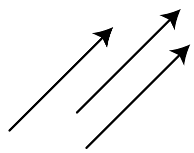
- Note the reverse order of multiplication in (15) and (19).
- (19) can only be applied if we know that both A and B are invertible.
- In general, $AB \neq BA$, even if A and B are both square. If $AB = BA$, then we say that A and B **commute**.
- For a general matrix A, we cannot say that $AB = AC$ yields $B = C$. (However, if we know that A is invertible, then we can multiply both sides of the equation $AB = AC$ to the left by A^{-1} and get $B = C$.)
- The equation $AB = 0$ does not necessarily yield $A = 0$ or $B = 0$. For example, take
$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

APPENDIX A

Vectors

CONTENTS

To a mathematician, a vector is the fundamental element of what is known as a vector space, supporting the operations of scaling, by elements known as scalars, and also supporting addition between vectors. When using vectors to describe physical quantities, like velocity, acceleration, and force, we can move away from this abstract definition, and stick with a more concrete notion. We can view them as arrows in space, of a particular length and denoting a particular direction, and we can think of the corresponding scalars as simply the real numbers. Practically speaking, a vector is simply a way of simultaneously storing and handling two pieces of information: a direction in space, and a magnitude or length.



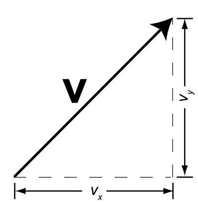
An arrow is a convenient way to draw a vector; since both length and direction are clearly indicated. A real number is a convenient way to represent a scalar, which when multiplied by a vector changes its length. To the left are three visual representations of identical vectors. They are identical, since they are all of the same length and the same direction, i.e. they are parallel to each other. Their location within the space is irrelevant.

In the study of physically based animation, we will initially be interested in vectors in two-dimensional (2D) and in three-dimensional (3D) space, whose elements are real numbers. But, we will see later that vectors can be defined in a space of any number of dimensions, with elements that may themselves be multidimensional.

Notationally, a vector is usually denoted by a lower-case letter, which has a line over it, like $\overline{v}$, or is printed in bold type, like $\mathbf{v}$. For hand written notes, the line is most convenient, but in printed form the bold form is more usual. Throughout these notes the form $\mathbf{v}$ is used.

A vector in 2D Euclidean space is defined by a pair of scalars arranged in a column, like

$$\mathbf{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}.$$



Examining the diagram to the right, we see that v_x denotes the horizontal extent or *component* of the vector, and v_y its vertical component. Note, that in a computer program this structure can be easily represented as a two-element array of floating point numbers, or a struct containing two floats. When working in 2D, the direction of the vector can be given by the slope $m = v_y/v_x$. Its magnitude, also called its *norm*, is written $\|\mathbf{v}\|$. By the Pythagorean Theorem, $\|\mathbf{v}\| = \sqrt{v_x^2 + v_y^2}$.

A vector in 3D space is defined by three scalars arranged in a column,

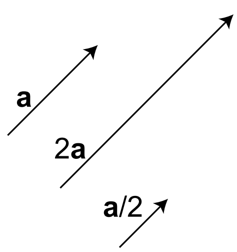
$$\mathbf{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix},$$

where v_x is the horizontal component, v_y the vertical component, and v_z the depth component. The norm of a 3D vector $\mathbf{v}$ is

$$\|\mathbf{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2}.$$

In 3D there is no simple equivalent to the slope. The direction of a 3D vector is often given in terms of its azimuth and elevation. But, for our purposes it will be best understood by its corresponding unit vector, which we will describe after first defining some key algebraic vector operations.

A.1 SCALING A VECTOR



Multiplication of a vector by a real number scalar leaves the vector's direction unchanged, but multiplies its magnitude by the scalar. Algebraically, we multiply each term of the vector by the scalar. For example

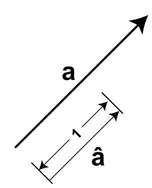
$$2\mathbf{a} = 2\begin{bmatrix} a_x \\ a_y \end{bmatrix} = \begin{bmatrix} 2a_x \\ 2a_y \end{bmatrix}.$$

Division by a scalar is the same as multiplication by the reciprocal of the scalar:

$$\mathbf{a}/2 = \begin{bmatrix} a_x/2 \\ a_y/2 \end{bmatrix}.$$

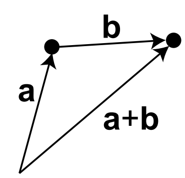
A.2 UNIT OR DIRECTION VECTORS

The direction of a vector is most easily described by a *unit vector*, also called a *direction vector*. A unit vector, for a particular vector, is parallel to that vector but of unit length. Therefore, it retains the direction, but not the norm of the parent vector. Throughout these notes the notation $\hat{\mathbf{v}}$ will be used to indicate a unit vector in the direction of parent vector $\mathbf{v}$. For example, the unit or direction vector corresponding with the 2D vector $\mathbf{a}$ would be



$$\hat{\mathbf{a}} = \begin{bmatrix} a_x/\|\mathbf{a}\| \\ a_y/\|\mathbf{a}\| \end{bmatrix} = \begin{bmatrix} \hat{a}_x \\ \hat{a}_y \end{bmatrix}.$$

A.3 VECTOR ADDITION

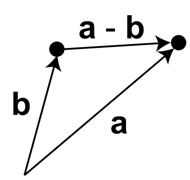


Addition of vectors can be expressed by a diagram. Placing the vectors end to end, the vector from the start of the first vector to the end of the second vector is the sum of the vectors. One way to think of this is that we start at the beginning of the first vector, travel along that vector to its end, and then travel from the start of the second vector to its end. An arrow constructed between the starting and ending points defines a new vector, which is the sum of the original vectors. Algebraically, this is equivalent to adding corresponding terms of the two vectors:

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_x \\ a_y \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \end{bmatrix} = \begin{bmatrix} a_x + b_x \\ a_y + b_y \end{bmatrix}.$$

We can think of this as again making a trip from the start of the first vector to the end of the second vector, but this time traveling first horizontally the distance $a_x + b_x$ and then vertically the distance $a_y + b_y$.

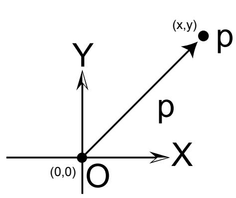
A.4 VECTOR SUBTRACTION



Subtraction of vectors can be shown in diagram form by placing the starting points of the two vectors together, and then constructing an arrow from the head of the second vector in the subtraction to the head of the first vector. Algebraically, we subtract corresponding terms:

$$\mathbf{a} - \mathbf{b} = \begin{bmatrix} a_x \\ a_y \end{bmatrix} - \begin{bmatrix} b_x \\ b_y \end{bmatrix} = \begin{bmatrix} a_x - b_x \\ a_y - b_y \end{bmatrix}.$$

A.5 POINTS AND VECTORS



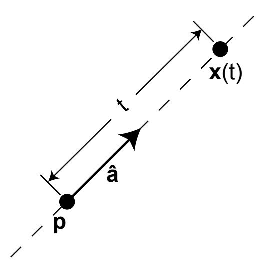
This leads us to the idea that points and vectors can be interchanged — almost. While vectors can exist anywhere in space, a point is always defined relative to the origin, O. Thus, we can say that a point, $p = (x, y)$, is defined by the origin, $O = (0, 0)$ and a vector, $\mathbf{p} = \begin{bmatrix} x \\ y \end{bmatrix}$, i.e.

$$p = O + \mathbf{p}.$$

Because the origin is assumed to be the point $(0, 0)$, points and vectors can be represented the same way, e.g. the point $(2, 3)$ can be represented as the vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$. This interchangeability can be very convenient in many cases, but can also lead to confusion. It is a good idea to make sure that when storing data, you clearly indicate which values are points, and which are vectors. As will be seen below, the homogeneous coordinates used to define transformations can help with this.

Equivalent to the above, we can write, $\mathbf{p} = p - O$, i.e. a vector defines the measure from the origin to a particular point in space. More generally, a vector can always be defined by the difference between any two points, p and q. The vector $\mathbf{v} = p - q$ represents the direction and distance from point q to point p. Conversely, the point q and the vector $\mathbf{v}$ define the point, $p = q + \mathbf{v}$, which is translated from q by the components of $\mathbf{v}$.

A.6 PARAMETRIC DEFINITION OF LINES AND RAYS



This leads us to a compact definition of a line in space, written in terms of a unit vector and a point. Let $\mathbf{p}$ be a known point (expressed in vector form) on the line being defined, and let $\hat{\mathbf{a}}$ be a unit vector whose direction is parallel to the desired line. Then, the locus of points on the line is the set of all points $\mathbf{x}$, satisfying

$$\mathbf{x}(t) = \mathbf{p} + t\hat{\mathbf{a}}.$$

The variable t is a real number, and is known as the line parameter. It measures the distance from the point $\mathbf{p}$ to the point $\mathbf{x}(t)$. If t is positive, the point $\mathbf{x}$ lies in the direction of the unit vector from point $\mathbf{p}$, and if t is negative, the point lies in the direction opposite to the unit vector.

The definition of a ray is identical to the definition of a line, except that the parameter t of a ray is limited to the positive real numbers. Thus, a ray can be interpreted as starting from the point $\mathbf{p}$, and traveling in the direction of $\hat{\mathbf{a}}$ a distance corresponding to t, as t goes from 0 to increasingly large positive values. On a ray, the point $\mathbf{p}$ is called the ray origin, $\hat{\mathbf{a}}$ the ray direction, and t the distance along the ray.

A.7 DOT OR INNER PRODUCT

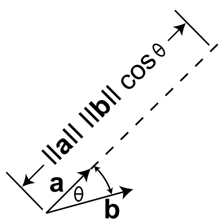
Vector-vector multiplication is not as easily defined as addition, subtraction and scalar multiplication. There are actually several vector products that can be defined. First, we will look at the *dot product* of two vectors, which is often called their *inner product*.

Defined algebraically, the dot product of two vectors is given by

$$\mathbf{a} \cdot \mathbf{b} = \begin{bmatrix} a_x \\ a_y \end{bmatrix} \cdot \begin{bmatrix} b_x \\ b_y \end{bmatrix} = a_x b_x + a_y b_y.$$

We multiply corresponding terms and add the result. The result *is not a vector*, but is in fact a scalar. This turns out to have many ramifications. The dot product is a *mighty* operation and has many uses in graphics!

A.7.1 Trigonometric interpretation of dot product



The dot product can be written in trigonometric form as

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta,$$

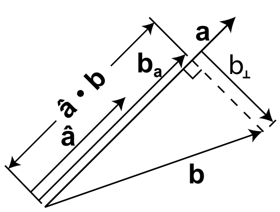
where θ is the smallest angle between the two vectors. Note, that this definition of θ applies in both 2D and 3D. Two nonparallel vectors always define a plane, and the angle θ is the angle between the vectors measured in that plane. Note that if both $\mathbf{a}$ and $\mathbf{b}$ are unit vectors, then $\|\mathbf{a}\|\|\mathbf{b}\| = 1$, and $\mathbf{a} \cdot \mathbf{b} = \cos\theta$. So, in general if you want to find the cosine of the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$, first compute the unit vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ in the directions of $\mathbf{a}$ and $\mathbf{b}$ then

$$\cos\theta = \hat{\mathbf{a}} \cdot \hat{\mathbf{b}}.$$

Other things to note about the trigonometric representation of dot product that follow directly from the cosine relationship are that

1. the dot product of *orthogonal* (perpendicular) vectors is zero, so if $\mathbf{a} \cdot \mathbf{b} = 0$, for vectors $\mathbf{a}$ and $\mathbf{b}$ with non-zero norms, we know that the vectors must be orthogonal,
2. the dot product of two vectors is positive if the magnitude of the smallest angle between the vectors is less than 90°, and negative if the magnitude of this angle exceeds 90°.

A.7.2 Geometric interpretation of dot product



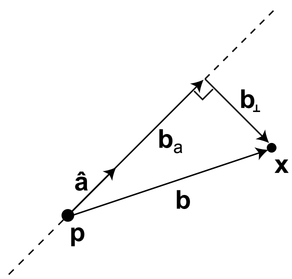
Another very useful interpretation of the dot product is that it can be used to compute the component of one vector in the direction parallel to another vector. For example, let $\hat{\mathbf{a}}$ be a unit vector in the direction of vector $\mathbf{a}$. Then the length of the projection of another vector $\mathbf{b}$ in the direction of vector $\mathbf{a}$ is $\hat{\mathbf{a}} \cdot \mathbf{b}$. You can think of this as the length of the shadow of vector $\mathbf{b}$ on vector $\mathbf{a}$. Therefore, the vector component of $\mathbf{b}$ in the direction of $\mathbf{a}$ is

$$\mathbf{b}_a = (\hat{\mathbf{a}} \cdot \mathbf{b})\hat{\mathbf{a}}.$$

So, $\mathbf{b}_a$ is parallel to $\mathbf{a}$ and has length equal to the projection of $\mathbf{b}$ onto $\mathbf{a}$. Note also that $\mathbf{b}_\perp = \mathbf{b} - \mathbf{b}_a$ will be the component of $\mathbf{b}$ perpendicular to vector $\mathbf{a}$.

The dot product has many uses in graphics that the following two examples will serve to illustrate.

A.7.3 Dot product example: The distance from a point to a line



Let us look at how dot product can be used to compute an important geometric quantity: the distance from a point to a line. We will use the parametric definition of a line, described above, specified by point $\mathbf{p}$ and a direction vector $\hat{\mathbf{a}}$. To compute the distance of an arbitrary point $\mathbf{x}$ from this line, first compute the vector $\mathbf{b} = \mathbf{x} - \mathbf{p}$, from the point $\mathbf{p}$ on the line to the point $\mathbf{x}$. The component of $\mathbf{b}$ in the direction of vector $\hat{\mathbf{a}}$ is

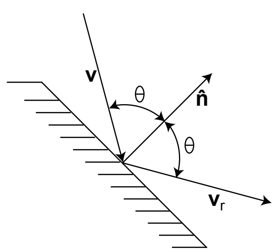
$$\mathbf{b}_a = (\hat{\mathbf{a}} \cdot \mathbf{b})\hat{\mathbf{a}}.$$

The component of $\mathbf{b}$ perpendicular to $\mathbf{a}$ is

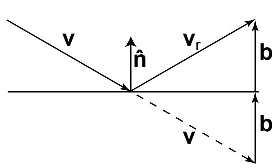
$$\mathbf{b}_\perp = \mathbf{b} - \mathbf{b}_a,$$

and the distance of point $\mathbf{x}$ from the line is simply $\|\mathbf{b}_\perp\|$.

A.7.4 Dot product example: Mirror reflection



Another very useful example of the use of dot product in geometric calculations is the computation of the mirror reflection from a surface. Assume that we have a flat mirror surface, whose *surface normal* is the unit vector $\hat{\mathbf{n}}$. The surface normal is defined to be a direction vector perpendicular to the surface. Since there are two such vectors at any point on a surface, the convention is to take the direction of the surface normal to be pointing in the "up" direction of the surface. For example, on a sphere it would point out of the sphere, and on a plane it would point in the direction considered to be the top of the plane. Now, we shine a light ray with direction $\mathbf{v}$ at the surface. The direction of the reflected ray will be given by $\mathbf{v}_r$. What must be true is that the angle θ between the normal $\hat{\mathbf{n}}$ and the light ray $\mathbf{v}$ should be the same as the angle between the reflected ray and the normal, and all three vectors $\mathbf{v}$, $\hat{\mathbf{n}}$, and $\mathbf{v}_r$ must lie in the same plane. Given these constraints, below is one way to calculate the light reflection ray $\mathbf{v}_r$.



To make the figure to the left, we first rotated the scene so everything is in a convenient orientation, with the surface normal $\hat{\mathbf{n}}$ pointing vertically, and the surface horizontal. Now, move vector $\mathbf{v}$ so that its tail is at the reflection point, as shown by the vector drawn with a dashed line in the figure. If $\mathbf{b}$ is the vector parallel to $\hat{\mathbf{n}}$ from the head of $\mathbf{v}$ to the surface, then by vector addition we have

$$\mathbf{v}_r = \mathbf{v} + 2\mathbf{b}.$$

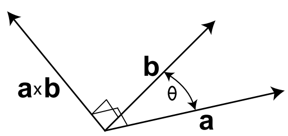
Now the vector $\mathbf{b}$ is just the negative of the component of $\mathbf{v}$ in the direction of $\hat{\mathbf{n}}$. So,

$$\mathbf{b} = -(\hat{\mathbf{n}} \cdot \mathbf{v})\hat{\mathbf{n}}.$$

Thus,

$$\mathbf{v}_r = \mathbf{v} - 2(\hat{\mathbf{n}} \cdot \mathbf{v})\hat{\mathbf{n}}.$$

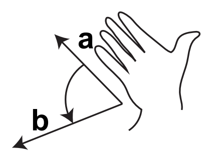
A.8 CROSS PRODUCT



The cross product $\mathbf{a} \times \mathbf{b}$ between two vectors $\mathbf{a}$ and $\mathbf{b}$ is a new vector perpendicular to the plane defined by the original two vectors. In other words, the cross product of two vectors is a vector that is perpendicular to both of the original vectors. The figure to the left illustrates the construction.

This notion of cross product does not make sense in 2D space, since it is not possible for a third 2D vector to be perpendicular to two (non parallel) 2D vectors. Thus, in graphics, the notion of cross product is reserved for working in 3D space.

Since there are two directions perpendicular to the plane formed by two vectors, we must have a convention to determine which of these two directions to use. In graphics, it is most common to use the *right hand rule*, and we use this convention throughout this text. The right-hand rule works as follows. Hold your right hand out flat, with the thumb out, aligning the fingers so they point in the direction of $\mathbf{a}$.



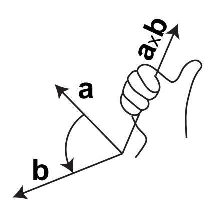
Now, rotate your hand so you can curl your fingers in the direction from vector $\mathbf{a}$ to vector $\mathbf{b}$. Your thumb will point in the direction of $\mathbf{a} \times \mathbf{b}$. If you reverse this, and first align your fingers with $\mathbf{b}$ and then curl them towards $\mathbf{a}$ you will see that you have to turn your hand upside down, reversing the direction in which your thumb is pointing. From this it should be apparent that $\mathbf{b} \times \mathbf{a} = -(\mathbf{a} \times \mathbf{b})$. In other words, the order of the operands in the cross product changes the polarity of the resulting cross product vector. The result is still perpendicular to both of the original vectors, but the direction is flipped.

A.8.1 Trigonometric interpretation of cross product

The magnitude of the cross product is given by

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\||\sin\theta|,$$

where θ is the small angle between vectors $\mathbf{a}$ and $\mathbf{b}$. Thus, if $\mathbf{a}$ and $\mathbf{b}$ are unit vectors, the magnitude of the cross product is the magnitude of $\sin\theta$.



Note, that the cross product of two parallel vectors will be the zero vector **0**. This is consistent with the geometric notion that the cross product produces a vector orthogonal to the original two vectors. If the original vectors are parallel, then there is no unique direction perpendicular to both vectors (i.e. there are infinitely many orthogonal vectors, all parallel to any plane perpendicular to either vector).

Algebraically, the cross product is defined as follows. If two vectors are defined

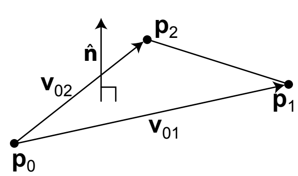
$$\mathbf{a} = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix},$$

then

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{bmatrix}.$$

The cross product has many uses in graphics, which the following two examples will serve to illustrate.

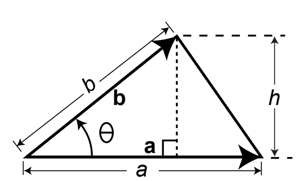
A.8.2 Cross product example: Finding surface normals



Suppose we have triangle $(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$, and we want to find the triangle's surface normal. We can do this easily by use of a cross product operation. First, define vectors along two of the triangle edges: $\mathbf{v}_{01} = \mathbf{p}_1 - \mathbf{p}_0$, and $\mathbf{v}_{02} = \mathbf{p}_2 - \mathbf{p}_0$. Then the cross product $\mathbf{v}_{01} \times \mathbf{v}_{02}$ is a vector perpendicular to both $\mathbf{v}_{01}$ and $\mathbf{v}_{02}$, and therefore perpendicular to the plane of the triangle. Scaling this vector to a unit vector yields the surface normal vector

$$\hat{\mathbf{n}} = (\mathbf{v}_{01} \times \mathbf{v}_{02})/\|\mathbf{v}_{01} \times \mathbf{v}_{02}\|.$$

A.8.3 Cross product example: Computing the area of a triangle



Another application of cross product to triangles uses the trigonometric definition of the magnitude of the cross product. Suppose we have a triangle, like the one shown to the right. If we know the lengths of sides a and b, and we know the angle θ between these sides, the area computation is straightforward. Relative to side a, the height of the triangle is given by $h = b \sin\theta$, and we know that the area of the triangle is $A = 1/2ah$, so we have $A = 1/2ab\sin\theta$. If we represent the sides of the triangle by vectors **a** and **b**, $a = \|\mathbf{a}\|$ and $b = \|\mathbf{b}\|$. Since the magnitude of the cross product $\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\||\sin\theta|$, it follows that

$$A = 1/2\|\mathbf{a} \times \mathbf{b}\|.$$

Numerical Analysis Lecture Notes

Peter J. Olver

5. Inner Products and Norms

The norm of a vector is a measure of its size. Besides the familiar Euclidean norm based on the dot product, there are a number of other important norms that are used in numerical analysis. In this section, we review the basic properties of inner products and norms.

5.1. Inner Products.

Some, but not all, norms are based on inner products. The most basic example is the familiar *dot product*

$$\langle \mathbf{v}\,;\mathbf{w}\,\rangle = \mathbf{v}\cdot\mathbf{w} = v_1\,w_1 + v_2\,w_2 + \;\cdots\; + v_n\,w_n = \sum_{i=1}^{n} v_i\,w_i, \tag{5.1}$$

between (column) vectors $\mathbf{v} = (\,v_1, v_2, \ldots, v_n\,)^T$, $\mathbf{w} = (\,w_1, w_2, \ldots, w_n\,)^T$, lying in the Euclidean space $\mathbb{R}^n$. A key observation is that the dot product (5.1) is equal to the matrix product

$$\mathbf{v}\cdot\mathbf{w} = \mathbf{v}^T\mathbf{w} = (\,v_1 \quad v_2 \quad \ldots \quad v_n\,)\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \tag{5.2}$$

between the row vector $\mathbf{v}^T$ and the column vector $\mathbf{w}$. The key fact is that the dot product of a vector with itself,

$$\mathbf{v}\cdot\mathbf{v} = v_1^2 + v_2^2 + \;\cdots\; + v_n^2,$$

is the sum of the squares of its entries, and hence, by the classical Pythagorean Theorem, equals the square of its length; see Figure 5.1. Consequently, the *Euclidean norm* or *length* of a vector is found by taking the square root:

$$\|\,\mathbf{v}\,\| \;=\; \sqrt{\mathbf{v}\cdot\mathbf{v}} \;=\; \sqrt{v_1^2 + v_2^2 + \;\cdots\; + v_n^2}\;. \tag{5.3}$$

Note that every nonzero vector $\mathbf{v} \neq \mathbf{0}$ has positive Euclidean norm, $\|\,\mathbf{v}\,\| > 0$, while only the zero vector has zero norm: $\|\,\mathbf{v}\,\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$. The elementary properties of dot product and Euclidean norm serve to inspire the abstract definition of more general inner products.

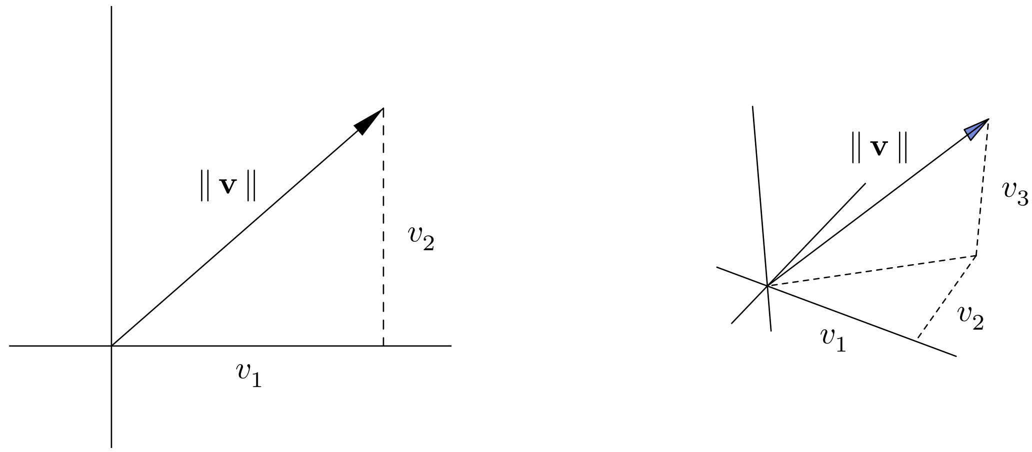


Figure 5.1. The Euclidean Norm in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 5.1. An *inner product* on the vector space $\mathbb{R}^n$ is a pairing that takes two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and produces a real number $\langle \mathbf{v} ; \mathbf{w} \rangle \in \mathbb{R}$. The inner product is required to satisfy the following three axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and scalars $c, d \in \mathbb{R}$.

(i) *Bilinearity*:

$$\begin{aligned} \langle c\,\mathbf{u} + d\,\mathbf{v} ; \mathbf{w} \rangle &= c\,\langle \mathbf{u} ; \mathbf{w} \rangle + d\,\langle \mathbf{v} ; \mathbf{w} \rangle, \\ \langle \mathbf{u} ; c\,\mathbf{v} + d\,\mathbf{w} \rangle &= c\,\langle \mathbf{u} ; \mathbf{v} \rangle + d\,\langle \mathbf{u} ; \mathbf{w} \rangle. \end{aligned} \tag{5.4}$$

(ii) *Symmetry*:

$$\langle \mathbf{v} ; \mathbf{w} \rangle = \langle \mathbf{w} ; \mathbf{v} \rangle. \tag{5.5}$$

(iii) *Positivity*:

$$\langle \mathbf{v} ; \mathbf{v} \rangle > 0 \qquad \text{whenever} \qquad \mathbf{v} \neq \mathbf{0}, \qquad \text{while} \qquad \langle \mathbf{0} ; \mathbf{0} \rangle = 0. \tag{5.6}$$

Given an inner product, the associated *norm* of a vector $\mathbf{v} \in V$ is defined as the positive square root of the inner product of the vector with itself:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v} ; \mathbf{v} \rangle}\,. \tag{5.7}$$

The positivity axiom implies that $\|\mathbf{v}\| \geq 0$ is real and non-negative, and equals 0 if and only if $\mathbf{v} = \mathbf{0}$ is the zero vector.

Example 5.2. While certainly the most common inner product on $\mathbb{R}^2$, the dot product

$$\mathbf{v} \cdot \mathbf{w} = v_1\,w_1 + v_2\,w_2$$

is by no means the only possibility. A simple example is provided by the *weighted inner product*

$$\langle \mathbf{v} ; \mathbf{w} \rangle = 2\,v_1\,w_1 + 5\,v_2\,w_2, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{5.8}$$

Let us verify that this formula does indeed define an inner product. The symmetry axiom (5.5) is immediate. Moreover,

$$\begin{aligned} \langle c\,\mathbf{u} + d\,\mathbf{v} ; \mathbf{w} \rangle &= 2\,(c u_1 + d\,v_1)\,w_1 + 5\,(c u_2 + d\,v_2)\,w_2 \\ &= c(2\,u_1\,w_1 + 5\,u_2\,w_2) + d\,(2\,v_1\,w_1 + 5\,v_2\,w_2) = c\,\langle \mathbf{u} ; \mathbf{w} \rangle + d\,\langle \mathbf{v} ; \mathbf{w} \rangle, \end{aligned}$$

which verifies the first bilinearity condition; the second follows by a very similar computation. Moreover, $\langle\, \mathbf{0}\,;\mathbf{0}\,\rangle = 0$, while

$$\langle\, \mathbf{v}\,;\mathbf{v}\,\rangle = 2\,v_1^2 + 5\,v_2^2 > 0 \qquad \text{whenever} \qquad \mathbf{v} \neq \mathbf{0},$$

since at least one of the summands is strictly positive. This establishes (5.8) as a legitimate inner product on $\mathbb{R}^2$. The associated *weighted norm* $\|\,\mathbf{v}\,\| = \sqrt{2\,v_1^2 + 5\,v_2^2}$ defines an alternative, "non-Pythagorean" notion of length of vectors and distance between points in the plane.

A less evident example of an inner product on $\mathbb{R}^2$ is provided by the expression

$$\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle = v_1\,w_1 - v_1\,w_2 - v_2\,w_1 + 4\,v_2\,w_2. \tag{5.9}$$

Bilinearity is verified in the same manner as before, and symmetry is immediate. Positivity is ensured by noticing that

$$\langle\, \mathbf{v}\,;\mathbf{v}\,\rangle = v_1^2 - 2\,v_1\,v_2 + 4\,v_2^2 = (v_1 - v_2)^2 + 3\,v_2^2 \;\geq\; 0$$

is always non-negative, and, moreover, is equal to zero if and only if $v_1 - v_2 = 0, v_2 = 0$, i.e., only when $\mathbf{v} = \mathbf{0}$. We conclude that (5.9) defines yet another inner product on $\mathbb{R}^2$, with associated norm

$$\|\,\mathbf{v}\,\| = \sqrt{\langle\, \mathbf{v}\,;\mathbf{v}\,\rangle} \;=\; \sqrt{v_1^2 - 2\,v_1\,v_2 + 4\,v_2^2}\,.$$

The second example (5.8) is a particular case of a general class of inner products.

Example 5.3. Let $c_1, \ldots, c_n > 0$ be a set of *positive* numbers. The corresponding *weighted inner product* and *weighted norm* on $\mathbb{R}^n$ are defined by

$$\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle \;=\; \sum_{i=1}^{n} c_i\,v_i\,w_i, \qquad\qquad \|\,\mathbf{v}\,\| \;=\; \sqrt{\langle\, \mathbf{v}\,;\mathbf{v}\,\rangle} \;=\; \sqrt{\sum_{i=1}^{n} c_i\,v_i^2}\,. \tag{5.10}$$

The numbers c_i are the *weights*. Observe that the larger the weight c_i, the more the i^{th} coordinate of $\mathbf{v}$ contributes to the norm. We can rewrite the weighted inner product in the useful vector form

$$\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle = \mathbf{v}^T C\,\mathbf{w}, \qquad \text{where} \qquad C = \begin{pmatrix} c_1 & 0 & 0 & \ldots & 0 \\ 0 & c_2 & 0 & \ldots & 0 \\ 0 & 0 & c_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & c_n \end{pmatrix} \tag{5.11}$$

is the diagonal *weight matrix*. Weighted norms are particularly relevant in statistics and data fitting, [**12**], where one wants to emphasize certain quantities and de-emphasize others; this is done by assigning appropriate weights to the different components of the data vector $\mathbf{v}$.

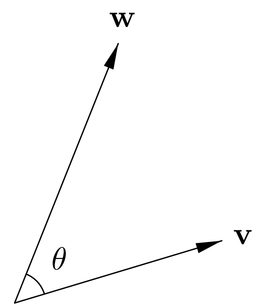


Figure 5.2. Angle Between Two Vectors.

5.2. Inequalities.

There are two absolutely fundamental inequalities that are valid for *any* inner product on any vector space. The first is inspired by the geometric interpretation of the dot product on Euclidean space in terms of the angle between vectors. It is named after two of the founders of modern analysis, Augustin Cauchy and Herman Schwarz, who established it in the case of the L^2 inner product on function space†. The more familiar triangle inequality, that the length of any side of a triangle is bounded by the sum of the lengths of the other two sides is, in fact, an immediate consequence of the Cauchy–Schwarz inequality, and hence also valid for any norm based on an inner product.

The Cauchy–Schwarz Inequality

In Euclidean geometry, the dot product between two vectors can be geometrically characterized by the equation

$$\mathbf{v} \cdot \mathbf{w} = \|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\|\,\cos\theta, \tag{5.12}$$

where θ measures the angle between the vectors $\mathbf{v}$ and $\mathbf{w}$, as drawn in Figure 5.2. Since

$$|\cos\theta\,| \leq 1,$$

the absolute value of the dot product is bounded by the product of the lengths of the vectors:

$$|\,\mathbf{v}\cdot\mathbf{w}\,| \;\leq\; \|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\|.$$

This is the simplest form of the general *Cauchy–Schwarz inequality*. We present a simple, algebraic proof that does not rely on the geometrical notions of length and angle and thus demonstrates its universal validity for *any* inner product.

Theorem 5.4. *Every inner product satisfies the Cauchy–Schwarz inequality*

$$|\,\langle\,\mathbf{v}\,;\mathbf{w}\,\rangle\,| \;\leq\; \|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\|, \qquad \textit{for all} \qquad \mathbf{v},\mathbf{w}\in V. \tag{5.13}$$

Here, $\|\,\mathbf{v}\,\|$ is the associated norm, while $|\cdot|$ denotes absolute value of real numbers. Equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel vectors.

† Russians also give credit for its discovery to their compatriot Viktor Bunyakovskii, and, indeed, some authors append his name to the inequality.

Proof: The case when $\mathbf{w} = \mathbf{0}$ is trivial, since both sides of (5.13) are equal to 0. Thus, we may suppose $\mathbf{w} \neq \mathbf{0}$. Let $t \in \mathbb{R}$ be an arbitrary scalar. Using the three inner product axioms, we have

$$0 \leq \| \mathbf{v} + t\,\mathbf{w} \|^2 = \langle \mathbf{v} + t\,\mathbf{w} \,; \mathbf{v} + t\,\mathbf{w} \rangle = \| \mathbf{v} \|^2 + 2t \langle \mathbf{v} \,; \mathbf{w} \rangle + t^2 \| \mathbf{w} \|^2, \tag{5.14}$$

with equality holding if and only if $\mathbf{v} = -\,t\,\mathbf{w}$ — which requires $\mathbf{v}$ and $\mathbf{w}$ to be parallel vectors. We fix $\mathbf{v}$ and $\mathbf{w}$, and consider the right hand side of (5.14) as a quadratic function,

$$0 \leq p(t) = a\,t^2 + 2\,b\,t + c, \qquad \text{where} \qquad a = \| \mathbf{w} \|^2, \qquad b = \langle \mathbf{v} \,; \mathbf{w} \rangle, \qquad c = \| \mathbf{v} \|^2,$$

of the scalar variable t. To get the maximum mileage out of the fact that $p(t) \geq 0$, let us look at where it assumes its minimum, which occurs when its derivative is zero:

$$p'(t) = 2\,a\,t + 2\,b = 0, \qquad \text{and so} \qquad t \;=\; -\,\frac{b}{a} \;=\; -\,\frac{\langle \mathbf{v} \,; \mathbf{w} \rangle}{\| \mathbf{w} \|^2}\,.$$

Substituting this particular value of t into (5.14), we find

$$0 \;\leq\; \| \mathbf{v} \|^2 - 2\,\frac{\langle \mathbf{v} \,; \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} + \frac{\langle \mathbf{v} \,; \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} \;=\; \| \mathbf{v} \|^2 - \frac{\langle \mathbf{v} \,; \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2}\,.$$

Rearranging this last inequality, we conclude that

$$\frac{\langle \mathbf{v} \,; \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} \;\leq\; \| \mathbf{v} \|^2, \qquad \text{or} \qquad \langle \mathbf{v} \,; \mathbf{w} \rangle^2 \;\leq\; \| \mathbf{v} \|^2 \, \| \mathbf{w} \|^2.$$

Also, as noted above, equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel. Taking the (positive) square root of both sides of the final inequality completes the proof of the Cauchy–Schwarz inequality (5.13). *Q.E.D.*

Given any inner product, we can use the quotient

$$\cos\theta = \frac{\langle \mathbf{v} \,; \mathbf{w} \rangle}{\| \mathbf{v} \| \, \| \mathbf{w} \|} \tag{5.15}$$

to define the "angle" between the vector space elements $\mathbf{v}, \mathbf{w} \in V$. The Cauchy–Schwarz inequality tells us that the ratio lies between -1 and $+1$, and hence the angle θ is well defined, and, in fact, unique if we restrict it to lie in the range $0 \leq \theta \leq \pi$.

For example, the vectors $\mathbf{v} = (1, 0, 1)^T$, $\mathbf{w} = (0, 1, 1)^T$ have dot product $\mathbf{v} \cdot \mathbf{w} = 1$ and norms $\| \mathbf{v} \| = \| \mathbf{w} \| = \sqrt{2}$. Hence the Euclidean angle between them is given by

$$\cos\theta = \frac{1}{\sqrt{2} \cdot \sqrt{2}} = \frac{1}{2}\,, \qquad \text{and so} \qquad \theta = \tfrac{1}{3}\,\pi = 1.0472\ldots\,.$$

On the other hand, if we adopt the weighted inner product $\langle \mathbf{v} \,; \mathbf{w} \rangle = v_1 w_1 + 2\,v_2 w_2 + 3\,v_3 w_3$, then $\mathbf{v} \cdot \mathbf{w} = 3$, $\| \mathbf{v} \| = 2$, $\| \mathbf{w} \| = \sqrt{5}$, and hence their "weighted" angle becomes

$$\cos\theta = \frac{3}{2\sqrt{5}} = .67082\ldots\,, \qquad \text{with} \qquad \theta = .835482\ldots\,.$$

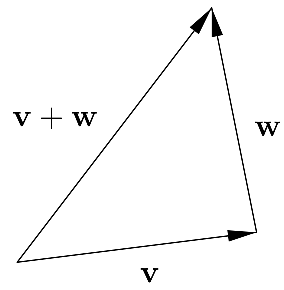


Figure 5.3. Triangle Inequality.

Thus, the measurement of angle (and length) is dependent upon the choice of an underlying inner product.

In Euclidean geometry, perpendicular vectors meet at a right angle, $\theta = \frac{1}{2}\pi$ or $\frac{3}{2}\pi$, with $\cos\theta = 0$. The angle formula (5.12) implies that the vectors $\mathbf{v}, \mathbf{w}$ are perpendicular if and only if their dot product vanishes: $\mathbf{v}\cdot\mathbf{w} = 0$. Perpendicularity is of interest in general inner product spaces, but, for historical reasons, has been given a more suggestive name.

Definition 5.5. Two elements $\mathbf{v}, \mathbf{w} \in V$ of an inner product space V are called *orthogonal* if their inner product vanishes: $\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle = 0$.

In particular, the zero element is orthogonal to everything: $\langle\, \mathbf{0}\,;\mathbf{v}\,\rangle = 0$ for all $\mathbf{v} \in V$. Orthogonality is a remarkably powerful tool in all applications of linear algebra, and often serves to dramatically simplify many computations.

The Triangle Inequality

The familiar triangle inequality states that the length of one side of a triangle is at most equal to the sum of the lengths of the other two sides. Referring to Figure 5.3, if the first two sides are represented by vectors $\mathbf{v}$ and $\mathbf{w}$, then the third corresponds to their sum $\mathbf{v}+\mathbf{w}$. The triangle inequality turns out to be an elementary consequence of the Cauchy–Schwarz inequality, and hence is valid in *any* inner product space.

Theorem 5.6. *The norm associated with an inner product satisfies the* triangle inequality

$$\|\,\mathbf{v}+\mathbf{w}\,\| \;\le\; \|\,\mathbf{v}\,\| + \|\,\mathbf{w}\,\| \qquad \text{for all} \qquad \mathbf{v}, \mathbf{w} \in V. \tag{5.16}$$

Equality holds if and only if $\mathbf{v}$ *and* $\mathbf{w}$ *are parallel vectors.*

Proof: We compute

$$\begin{aligned}
\|\,\mathbf{v}+\mathbf{w}\,\|^2 &= \langle\, \mathbf{v}+\mathbf{w}\,;\mathbf{v}+\mathbf{w}\,\rangle = \|\,\mathbf{v}\,\|^2 + 2\,\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle + \|\,\mathbf{w}\,\|^2 \\
&\le \|\,\mathbf{v}\,\|^2 + 2\,|\,\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle\,| + \|\,\mathbf{w}\,\|^2 \le \|\,\mathbf{v}\,\|^2 + 2\,\|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\| + \|\,\mathbf{w}\,\|^2 \\
&= \big(\,\|\,\mathbf{v}\,\| + \|\,\mathbf{w}\,\|\,\big)^2,
\end{aligned}$$

where the middle inequality follows from Cauchy–Schwarz. Taking square roots of both sides and using positivity completes the proof. *Q.E.D.*

Example 5.7. The vectors $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$ sum to $\mathbf{v} + \mathbf{w} = \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}$. Their Euclidean norms are $\| \mathbf{v} \| = \sqrt{6}$ and $\| \mathbf{w} \| = \sqrt{13}$, while $\| \mathbf{v} + \mathbf{w} \| = \sqrt{17}$. The triangle inequality (5.16) in this case says $\sqrt{17} \leq \sqrt{6} + \sqrt{13}$, which is valid.

5.3. Norms.

Every inner product gives rise to a norm that can be used to measure the magnitude or length of the elements of the underlying vector space. However, not every norm that is used in analysis and applications arises from an inner product. To define a general norm, we will extract those properties that do not directly rely on the inner product structure.

Definition 5.8. A *norm* on the vector space $\mathbb{R}^n$ assigns a real number $\| \mathbf{v} \|$ to each vector $\mathbf{v} \in V$, subject to the following axioms for every $\mathbf{v}, \mathbf{w} \in V$, and $c \in \mathbb{R}$.

(i) *Positivity*: $\| \mathbf{v} \| \geq 0$, with $\| \mathbf{v} \| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(ii) *Homogeneity*: $\| c\, \mathbf{v} \| = | c | \, \| \mathbf{v} \|$.

(iii) *Triangle inequality*: $\| \mathbf{v} + \mathbf{w} \| \leq \| \mathbf{v} \| + \| \mathbf{w} \|$.

As we now know, every inner product gives rise to a norm. Indeed, positivity of the norm is one of the inner product axioms. The homogeneity property follows since

$$\| c\, \mathbf{v} \| = \sqrt{\langle c\, \mathbf{v} \,; c\, \mathbf{v} \rangle} = \sqrt{c^2 \langle \mathbf{v} \,; \mathbf{v} \rangle} = | c | \sqrt{\langle \mathbf{v} \,; \mathbf{v} \rangle} = | c | \, \| \mathbf{v} \|.$$

Finally, the triangle inequality for an inner product norm was established in Theorem 5.6. Let us introduce some of the principal examples of norms that do not come from inner products.

The 1–*norm* of a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T \in \mathbb{R}^n$ is defined as the sum of the absolute values of its entries:

$$\| \mathbf{v} \|_1 = | v_1 | + | v_2 | + \cdots + | v_n |. \tag{5.17}$$

The *max* or ∞–*norm* is equal to its maximal entry (in absolute value):

$$\| \mathbf{v} \|_\infty = \max \{ \, | v_1 |, | v_2 |, \ \ldots \ , | v_n | \, \}. \tag{5.18}$$

Verification of the positivity and homogeneity properties for these two norms is straightforward; the triangle inequality is a direct consequence of the elementary inequality

$$| a + b | \leq | a | + | b |, \qquad a, b \in \mathbb{R},$$

for absolute values.

The Euclidean norm, 1–norm, and ∞–norm on $\mathbb{R}^n$ are just three representatives of the general p–*norm*

$$\| \mathbf{v} \|_p = \sqrt[p]{\sum_{i=1}^{n} | v_i |^p} \; . \tag{5.19}$$

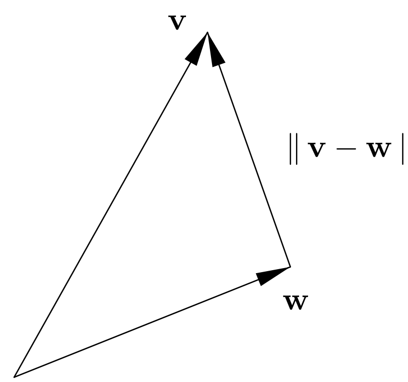


Figure 5.4. Distance Between Vectors.

This quantity defines a norm for any $1 \leq p < \infty$. The ∞–norm is a limiting case of (5.19) as $p \to \infty$. Note that the Euclidean norm (5.3) is the 2–norm, and is often designated as such; it is the only p–norm which comes from an inner product. The positivity and homogeneity properties of the p–norm are not hard to establish. The triangle inequality, however, is not trivial; in detail, it reads

$$\sqrt[p]{\sum_{i=1}^{n} |\, v_i + w_i \,|^p} \;\leq\; \sqrt[p]{\sum_{i=1}^{n} |\, v_i \,|^p} \;+\; \sqrt[p]{\sum_{i=1}^{n} |\, w_i \,|^p}\,, \tag{5.20}$$

and is known as *Minkowski's inequality*. A complete proof can be found in [**31**].

Every norm defines a *distance* between vector space elements, namely

$$d(\mathbf{v},\mathbf{w}) = \| \mathbf{v} - \mathbf{w} \|. \tag{5.21}$$

For the standard dot product norm, we recover the usual notion of distance between points in Euclidean space. Other types of norms produce alternative (and sometimes quite useful) notions of distance that are, nevertheless, subject to all the familiar properties:

(*a*) *Symmetry*: $d(\mathbf{v},\mathbf{w}) = d(\mathbf{w},\mathbf{v})$;
(*b*) *Positivity*: $d(\mathbf{v},\mathbf{w}) = 0$ if and only if $\mathbf{v} = \mathbf{w}$;
(*c*) *Triangle Inequality*: $d(\mathbf{v},\mathbf{w}) \leq d(\mathbf{v},\mathbf{z}) + d(\mathbf{z},\mathbf{w})$.

Equivalence of Norms

While there are many different types of norms on $\mathbb{R}^n$, in a certain sense, they are all more or less equivalent†. "Equivalence" does not mean that they assume the same value, but rather that they are always close to one another, and so, for many analytical purposes, may be used interchangeably. As a consequence, we may be able to simplify the analysis of a problem by choosing a suitably adapted norm.

† This statement remains valid in any finite-dimensional vector space, but is *not* correct in infinite-dimensional function spaces.

Theorem 5.9. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be any two norms on $\mathbb{R}^n$. Then there exist positive constants $c^\star, C^\star > 0$ such that*

$$c^\star \, \| \mathbf{v} \|_1 \le \| \mathbf{v} \|_2 \le C^\star \, \| \mathbf{v} \|_1 \qquad \textit{for every} \qquad \mathbf{v} \in \mathbb{R}^n. \tag{5.22}$$

Proof: We just sketch the basic idea, leaving the details to a more rigorous real analysis course, cf. [**11**; §7.6]. We begin by noting that a norm defines a continuous real-valued function $f(\mathbf{v}) = \| \mathbf{v} \|$ on $\mathbb{R}^n$. (Continuity is, in fact, a consequence of the triangle inequality.) Let $S_1 = \left\{ \| \mathbf{u} \|_1 = 1 \right\}$ denote the unit sphere of the first norm. Any continuous function defined on a compact set achieves both a maximum and a minimum value. Thus, restricting the second norm function to the unit sphere S_1 of the first norm, we can set

$$c^\star = \min \{ \| \mathbf{u} \|_2 \mid \mathbf{u} \in S_1 \}, \qquad C^\star = \max \{ \| \mathbf{u} \|_2 \mid \mathbf{u} \in S_1 \}. \tag{5.23}$$

Moreover, $0 < c^\star \le C^\star < \infty$, with equality holding if and only if the norms are the same. The minimum and maximum (5.23) will serve as the constants in the desired inequalities (5.22). Indeed, by definition,

$$c^\star \le \| \mathbf{u} \|_2 \le C^\star \qquad \text{when} \qquad \| \mathbf{u} \|_1 = 1, \tag{5.24}$$

which proves that (5.22) is valid for all unit vectors $\mathbf{v} = \mathbf{u} \in S_1$. To prove the inequalities in general, assume $\mathbf{v} \ne \mathbf{0}$. (The case $\mathbf{v} = \mathbf{0}$ is trivial.) The homogeneity property of the norm implies that $\mathbf{u} = \mathbf{v}/\| \mathbf{v} \|_1 \in S_1$ is a unit vector in the first norm: $\| \mathbf{u} \|_1 = \| \mathbf{v} \| / \| \mathbf{v} \|_1 = 1$. Moreover, $\| \mathbf{u} \|_2 = \| \mathbf{v} \|_2 / \| \mathbf{v} \|_1$. Substituting into (5.24) and clearing denominators completes the proof of (5.22). *Q.E.D.*

Example 5.10. For example, consider the Euclidean norm $\|\cdot\|_2$ and the max norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$. The bounding constants are found by minimizing and maximizing $\| \mathbf{u} \|_\infty = \max\{ \, |u_1|, \ldots, |u_n| \, \}$ over all unit vectors $\| \mathbf{u} \|_2 = 1$ on the (round) unit sphere. The maximal value is achieved at the poles $\pm \mathbf{e}_k$, with $\| \pm \mathbf{e}_k \|_\infty = C^\star = 1$ The minimal value is attained at the points $\left(\pm \frac{1}{\sqrt{n}}, \ldots, \pm \frac{1}{\sqrt{n}} \right)$, whereby $c^\star = \frac{1}{\sqrt{n}}$. Therefore,

$$\frac{1}{\sqrt{n}} \, \| \mathbf{v} \|_2 \le \| \mathbf{v} \|_\infty \le \| \mathbf{v} \|_2. \tag{5.25}$$

We can interpret these inequalities as follows. Suppose $\mathbf{v}$ is a vector lying on the unit sphere in the Euclidean norm, so $\| \mathbf{v} \|_2 = 1$. Then (5.25) tells us that its ∞ norm is bounded from above and below by $\frac{1}{\sqrt{n}} \le \| \mathbf{v} \|_\infty \le 1$. Therefore, the Euclidean unit sphere sits inside the ∞ norm unit sphere and outside the ∞ norm sphere of radius $\frac{1}{\sqrt{n}}$.

11.3: The Dot Product

Learning Objectives

- Calculate the dot product of two given vectors.
- Determine whether two given vectors are perpendicular.
- Find the direction cosines of a given vector.
- Explain what is meant by the vector projection of one vector onto another vector, and describe how to compute it.
- Calculate the work done by a given force.

If we apply a force to an object so that the object moves, we say that work is done by the force. Previously, we looked at a constant force and we assumed the force was applied in the direction of motion of the object. Under those conditions, work can be expressed as the product of the force acting on an object and the distance the object moves. In this chapter, however, we have seen that both force and the motion of an object can be represented by vectors.

In this section, we develop an operation called the dot product, which allows us to calculate work in the case when the force vector and the motion vector have different directions. The dot product essentially tells us how much of the force vector is applied in the direction of the motion vector. The dot product can also help us measure the angle formed by a pair of vectors and the position of a vector relative to the coordinate axes. It even provides a simple test to determine whether two vectors meet at a right angle.

The Dot Product and Its Properties

We have already learned how to add and subtract vectors. In this chapter, we investigate two types of vector multiplication. The first type of vector multiplication is called the dot product, based on the notation we use for it, and it is defined as follows:

Definition: dot product

The *dot product* of vectors $\vec{\mathbf{u}} = \langle u_1, u_2, u_3 \rangle$ and $\vec{\mathbf{v}} = \langle v_1, v_2, v_3 \rangle$ is given by the sum of the products of the components

$$\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = u_1v_1 + u_2v_2 + u_3v_3. \tag{11.3.1}$$

Note that if u and v are two-dimensional vectors, we calculate the dot product in a similar fashion. Thus, if $\vec{\mathbf{u}} = \langle u_1, u_2 \rangle$ and $\vec{\mathbf{v}} = \langle v_1, v_2 \rangle$, then

$$\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = u_1v_1 + u_2v_2. \tag{11.3.2}$$

When two vectors are combined under addition or subtraction, the result is a vector. When two vectors are combined using the dot product, the result is a scalar. For this reason, the dot product is often called the *scalar product*. It may also be called the *inner product*.

Example 11.3.1: Calculating Dot Products

a. Find the dot product of $\vec{\mathbf{u}} = \langle 3, 5, 2 \rangle$ and $\vec{\mathbf{v}} = \langle -1, 3, 0 \rangle$.
b. Find the scalar product of $\vec{\mathbf{p}} = 10\hat{\mathbf{i}} - 4\hat{\mathbf{j}} + 7\hat{\mathbf{k}}$ and $\vec{\mathbf{q}} = -2\hat{\mathbf{i}} + \hat{\mathbf{j}} + 6\hat{\mathbf{k}}$.

Solution:

a. Substitute the vector components into the formula for the dot product:

$$\begin{aligned} \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} &= u_1v_1 + u_2v_2 + u_3v_3 \\ &= 3(-1) + 5(3) + 2(0) \\ &= -3 + 15 + 0 \\ &= 12. \end{aligned}$$

b. The calculation is the same if the vectors are written using standard unit vectors. We still have three components for each vector to substitute into the formula for the dot product:

$$\begin{aligned}\vec{\mathbf{p}}\cdot\vec{\mathbf{q}} &= u_1v_1+u_2v_2+u_3v_3\\ &=10(-2)+(-4)(1)+(7)(6)\\ &=-20-4+42\\ &=18.\end{aligned}$$

Exercise 11.3.1

Find $\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}$, where $\vec{\mathbf{u}}=\langle 2,9,-1\rangle$ and $\vec{\mathbf{v}}=\langle -3,1,-4\rangle$.

Like vector addition and subtraction, the dot product has several algebraic properties. We prove three of these properties and leave the rest as exercises.

Properties of the Dot Product

Let $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, and $\vec{\mathbf{w}}$ be vectors, and let c be a scalar.

i. *Commutative property*

$$\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}=\vec{\mathbf{v}}\cdot\vec{\mathbf{u}} \tag{11.3.3}$$

ii. *Distributive property*

$$\vec{\mathbf{u}}\cdot(\vec{\mathbf{v}}+\vec{\mathbf{w}})=\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}+\vec{\mathbf{u}}\cdot\vec{\mathbf{w}} \tag{11.3.4}$$

iii. *Associative property*

$$c(\vec{\mathbf{u}}\cdot\vec{\mathbf{v}})=(c\vec{\mathbf{u}})\cdot\vec{\mathbf{v}}=\vec{\mathbf{u}}\cdot(c\vec{\mathbf{v}}) \tag{11.3.5}$$

iv. *Property of magnitude*

$$\vec{\mathbf{v}}\cdot\vec{\mathbf{v}}=\|\vec{\mathbf{v}}\|^2 \tag{11.3.6}$$

Proof

Let $\vec{\mathbf{u}}=\langle u_1,u_2,u_3\rangle$ and $\vec{\mathbf{v}}=\langle v_1,v_2,v_3\rangle$. Then

$$\begin{aligned}\vec{\mathbf{u}}\cdot\vec{\mathbf{v}} &=\langle u_1,u_2,u_3\rangle\cdot\langle v_1,v_2,v_3\rangle\\ &=u_1v_1+u_2v_2+u_3v_3\\ &=v_1u_1+v_2u_2+v_3u_3\\ &=\langle v_1,v_2,v_3\rangle\cdot\langle u_1,u_2,u_3\rangle\\ &=\vec{\mathbf{v}}\cdot\vec{\mathbf{u}}.\end{aligned}$$

The associative property looks like the associative property for real-number multiplication, but pay close attention to the difference between scalar and vector objects:

$$\begin{aligned}c(\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}) &=c(u_1v_1+u_2v_2+u_3v_3)\\ &=c(u_1v_1)+c(u_2v_2)+c(u_3v_3)\\ &=(cu_1)v_1+(cu_2)v_2+(cu_3)v_3\\ &=\langle cu_1,cu_2,cu_3\rangle\cdot\langle v_1,v_2,v_3\rangle\\ &=c\langle u_1,u_2,u_3\rangle\cdot\langle v_1,v_2,v_3\rangle\\ &=(c\vec{\mathbf{u}})\cdot\vec{\mathbf{v}}.\end{aligned}$$

The proof that $c(\vec{\mathbf{u}}\cdot\vec{\mathbf{v}})=\vec{\mathbf{u}}\cdot(c\vec{\mathbf{v}})$ is similar.

The fourth property shows the relationship between the magnitude of a vector and its dot product with itself:

$$\begin{aligned}\vec{\mathbf{v}}\cdot\vec{\mathbf{v}} &=\langle v_1,v_2,v_3\rangle\cdot\langle v_1,v_2,v_3\rangle\\ &=(v_1)^2+(v_2)^2+(v_3)^2\\ &=\left[\sqrt{(v_1)^2+(v_2)^2+(v_3)^2}\right]^2\\ &=\|\vec{\mathbf{v}}\|^2.\end{aligned}$$

□

Note that by property iv. we have $\vec{\mathbf{0}}\cdot\vec{\mathbf{v}}=0.$ Also by property iv. if $\vec{\mathbf{v}}\cdot\vec{\mathbf{v}}=0,$ then $\vec{\mathbf{v}}=\vec{\mathbf{0}}.$

Example 11.3.2: Using Properties of the Dot Product

Let $\vec{\mathbf{a}}=\langle 1,2,-3\rangle$, $\vec{\mathbf{b}}=\langle 0,2,4\rangle$, and $\vec{\mathbf{c}}=\langle 5,-1,3\rangle$.

Find each of the following products.

a. $(\vec{\mathbf{a}}\cdot\vec{\mathbf{b}})\vec{\mathbf{c}}$
b. $\vec{\mathbf{a}}\cdot(2\vec{\mathbf{c}})$
c. $\|\vec{\mathbf{b}}\|^2$

Solution

a. Note that this expression asks for the scalar multiple of $\vec{\mathbf{c}}$ by $\vec{\mathbf{a}}\cdot\vec{\mathbf{b}}$:

$$\begin{aligned}(\vec{\mathbf{a}}\cdot\vec{\mathbf{b}})\vec{\mathbf{c}} &=(\langle 1,2,-3\rangle\cdot\langle 0,2,4\rangle)\langle 5,-1,3\rangle\\ &=(1(0)+2(2)+(-3)(4))\langle 5,-1,3\rangle\\ &=-8\langle 5,-1,3\rangle\\ &=\langle -40,8,-24\rangle.\end{aligned}$$

b. This expression is a dot product of vector $\vec{\mathbf{a}}$ and scalar multiple $2\vec{\mathbf{c}}$:

$$\begin{aligned}\vec{\mathbf{a}}\cdot(2\vec{\mathbf{c}}) &=2(\vec{\mathbf{a}}\cdot\vec{\mathbf{c}})\\ &=2(\langle 1,2,-3\rangle\cdot\langle 5,-1,3\rangle)\\ &=2(1(5)+2(-1)+(-3)(3))\\ &=2(-6)=-12.\end{aligned}$$

c. Simplifying this expression is a straightforward application of the dot product:

$$\begin{aligned}\|\vec{\mathbf{b}}\|^2 &=\vec{\mathbf{b}}\cdot\vec{\mathbf{b}}\\ &=\langle 0,2,4\rangle\cdot\langle 0,2,4\rangle\\ &=0^2+2^2+4^2\\ &=0+4+16\\ &=20.\end{aligned}$$

Exercise 11.3.2

Find the following products for $\vec{\mathbf{p}}=\langle 7,0,2\rangle$, $\vec{\mathbf{q}}=\langle -2,2,-2\rangle$, and $\vec{\mathbf{r}}=\langle 0,2,-3\rangle$.

a. $(\vec{\mathbf{r}}\cdot\vec{\mathbf{p}})\vec{\mathbf{q}}$

b. $\|\vec{\mathbf{p}}\|^2$

Using the Dot Product to Find the Angle between Two Vectors

When two nonzero vectors are placed in standard position, whether in two dimensions or three dimensions, they form an angle between them (Figure $11.3.1$). The dot product provides a way to find the measure of this angle. This property is a result of the fact that we can express the dot product in terms of the cosine of the angle formed by two vectors.

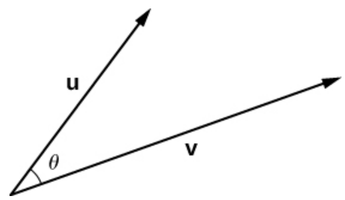


Figure 11.3.1: Let θ be the angle between two nonzero vectors $\vec{\mathbf{u}}$ *and* $\vec{\mathbf{v}}$ *such that* $0 \leq \theta \leq \pi$.

Evaluating a Dot Product

The dot product of two vectors is the product of the magnitude of each vector and the cosine of the angle between them:

$$\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = \|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta. \tag{11.3.7}$$

Proof

Place vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ in standard position and consider the vector $\vec{\mathbf{v}} - \vec{\mathbf{u}}$ (Figure $11.3.2$). These three vectors form a triangle with side lengths $\|\vec{\mathbf{u}}\|$, $\|\vec{\mathbf{v}}\|$, and $\|\vec{\mathbf{v}} - \vec{\mathbf{u}}\|$.

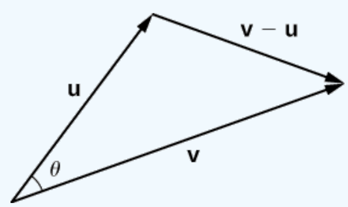


Figure 11.3.2: The lengths of the sides of the triangle are given by the magnitudes of the vectors that form the triangle.

Recall from trigonometry that the law of cosines describes the relationship among the side lengths of the triangle and the angle θ. Applying the law of cosines here gives

$$\|\vec{\mathbf{v}} - \vec{\mathbf{u}}\|^2 = \|\vec{\mathbf{u}}\|^2 + \|\vec{\mathbf{v}}\|^2 - 2\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta. \tag{11.3.8}$$

The dot product provides a way to rewrite the left side of Equation 11.3.8:

$$\begin{aligned} \|\vec{\mathbf{v}} - \vec{\mathbf{u}}\|^2 &= (\vec{\mathbf{v}} - \vec{\mathbf{u}}) \cdot (\vec{\mathbf{v}} - \vec{\mathbf{u}}) \\ &= (\vec{\mathbf{v}} - \vec{\mathbf{u}}) \cdot \vec{\mathbf{v}} - (\vec{\mathbf{v}} - \vec{\mathbf{u}}) \cdot \vec{\mathbf{u}} \\ &= \vec{\mathbf{v}} \cdot \vec{\mathbf{v}} - \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} - \vec{\mathbf{v}} \cdot \vec{\mathbf{u}} + \vec{\mathbf{u}} \cdot \vec{\mathbf{u}} \\ &= \vec{\mathbf{v}} \cdot \vec{\mathbf{v}} - \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} - \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} + \vec{\mathbf{u}} \cdot \vec{\mathbf{u}} \\ &= \|\vec{\mathbf{v}}\|^2 - 2\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} + \|\vec{\mathbf{u}}\|^2. \end{aligned}$$

Substituting into the law of cosines yields

$$\begin{aligned} \|\vec{\mathbf{v}} - \vec{\mathbf{u}}\|^2 &= \|\vec{\mathbf{u}}\|^2 + \|\vec{\mathbf{v}}\|^2 - 2\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta \\ \|\vec{\mathbf{v}}\|^2 - 2\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} + \|\vec{\mathbf{u}}\|^2 &= \|\vec{\mathbf{u}}\|^2 + \|\vec{\mathbf{v}}\|^2 - 2\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta \\ -2\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} &= -2\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta \\ \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} &= \|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\| \cos\theta. \end{aligned}$$

□

We can use the form of the dot product in Equation 11.3.7 to find the measure of the angle between two nonzero vectors by rearranging Equation 11.3.7 to solve for the cosine of the angle:

$$\cos\theta = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|}. \tag{11.3.9}$$

Using this equation, we can find the cosine of the angle between two nonzero vectors. Since we are considering the smallest angle between the vectors, we assume $0^\circ \leq \theta \leq 180^\circ$ (or $0 \leq \theta \leq \pi$ if we are working in radians). The inverse cosine is unique over this range, so we are then able to determine the measure of the angle θ.

Example 11.3.3: Finding the Angle between Two Vectors

Find the measure of the angle between each pair of vectors.

a. $\hat{\mathbf{i}}+\hat{\mathbf{j}}+\hat{\mathbf{k}}$ and $2\hat{\mathbf{i}}-\hat{\mathbf{j}}-3\hat{\mathbf{k}}$
b. $\langle 2,5,6\rangle$ and $\langle -2,-4,4\rangle$

Solution

a. To find the cosine of the angle formed by the two vectors, substitute the components of the vectors into Equation 11.3.9:

$$\begin{aligned}\cos\theta &= \frac{(\hat{\mathbf{i}}+\hat{\mathbf{j}}+\hat{\mathbf{k}})\cdot(2\hat{\mathbf{i}}-\hat{\mathbf{j}}-3\hat{\mathbf{k}})}{\left\|\hat{\mathbf{i}}+\hat{\mathbf{j}}+\hat{\mathbf{k}}\right\|\cdot\left\|2\hat{\mathbf{i}}-\hat{\mathbf{j}}-3\hat{\mathbf{k}}\right\|}\\ &= \frac{1(2)+(1)(-1)+(1)(-3)}{\sqrt{1^2+1^2+1^2}\sqrt{2^2+(-1)^2+(-3)^2}}\\ &= \frac{-2}{\sqrt{3}\sqrt{14}} = \frac{-2}{\sqrt{42}}.\end{aligned}$$

Therefore, $\theta = \arccos\dfrac{-2}{\sqrt{42}}$ rad.

b. Start by finding the value of the cosine of the angle between the vectors:

$$\begin{aligned}\cos\theta &= \frac{\langle 2,5,6\rangle\cdot\langle -2,-4,4\rangle}{\|\langle 2,5,6\rangle\|\cdot\|\langle -2,-4,4\rangle\|}\\ &= \frac{2(-2)+(5)(-4)+(6)(4)}{\sqrt{2^2+5^2+6^2}\sqrt{(-2)^2+(-4)^2+4^2}}\\ &= \frac{0}{\sqrt{65}\sqrt{36}} = 0.\end{aligned}$$

Now, $\cos\theta = 0$ and $0 \leq \theta \leq \pi$, so $\theta = \pi/2$.

Exercise 11.3.3

Find the measure of the angle, in radians, formed by vectors $\vec{\mathbf{a}} = \langle 1,2,0\rangle$ and $\vec{\mathbf{b}} = \langle 2,4,1\rangle$. Round to the nearest hundredth.

The angle between two vectors can be acute $(0 < \cos\theta < 1)$, obtuse $(-1 < \cos\theta < 0)$, or straight $(\cos\theta = -1)$. If $\cos\theta = 1$, then both vectors have the same direction. If $\cos\theta = 0$, then the vectors, when placed in standard position, form a right angle (Figure 11.3.3). We can formalize this result into a theorem regarding orthogonal (perpendicular) vectors.

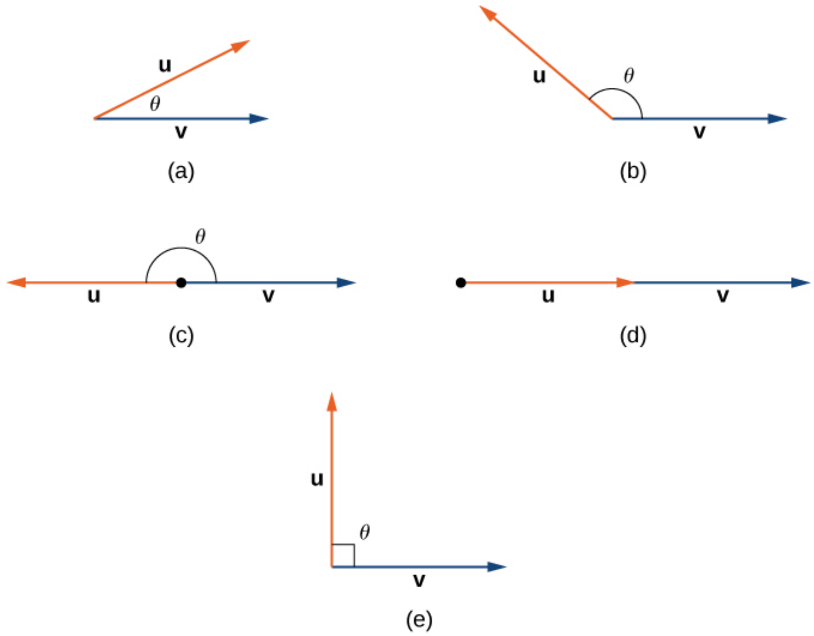


Figure 11.3.3: (a) An acute angle has $0<\cos\theta<1$. (b) An obtuse angle has $-1<\cos\theta<0$. (c) A straight line has $\cos\theta=-1$. (d) If the vectors have the same direction, $\cos\theta=1$. (e) If the vectors are orthogonal (perpendicular), $\cos\theta=0$.

Orthogonal Vectors

The nonzero vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ are **orthogonal vectors** if and only if $\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}=0$.

Proof

Let $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ be nonzero vectors, and let θ denote the angle between them. First, assume $\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}=0$. Then

$$\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|\cos\theta=0. \tag{11.3.10}$$

However, $\|\vec{\mathbf{u}}\|\neq 0$ and $\|\vec{\mathbf{v}}\|\neq 0$, so we must have $\cos\theta=0$. Hence, $\theta=90^\circ$, and the vectors are orthogonal.

Now assume $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ are orthogonal. Then $\theta=90^\circ$ and we have

$$\begin{aligned}\vec{\mathbf{u}}\cdot\vec{\mathbf{v}} &=\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|\cos\theta\\ &=\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|\cos 90^\circ\\ &=\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|(0)\\ &=0.\end{aligned}$$

□

The terms **orthogonal, perpendicular,** and **normal** each indicate that mathematical objects are intersecting at right angles. The use of each term is determined mainly by its context. We say that vectors are orthogonal and lines are perpendicular. The term *normal* is used most often when measuring the angle made with a plane or other surface.

Example 11.3.4: Identifying Orthogonal Vectors

Determine whether $\vec{\mathbf{p}}=\langle 1,0,5\rangle$ and $\vec{\mathbf{q}}=\langle 10,3,-2\rangle$ are orthogonal vectors.

Solution

Using the definition, we need only check the dot product of the vectors:

$$\vec{\mathbf{p}}\cdot\vec{\mathbf{q}}=1(10)+(0)(3)+(5)(-2)=10+0-10=0.$$

Because $\vec{\mathbf{p}}\cdot\vec{\mathbf{q}}=0,$ the vectors are orthogonal (Figure 11.3.4).

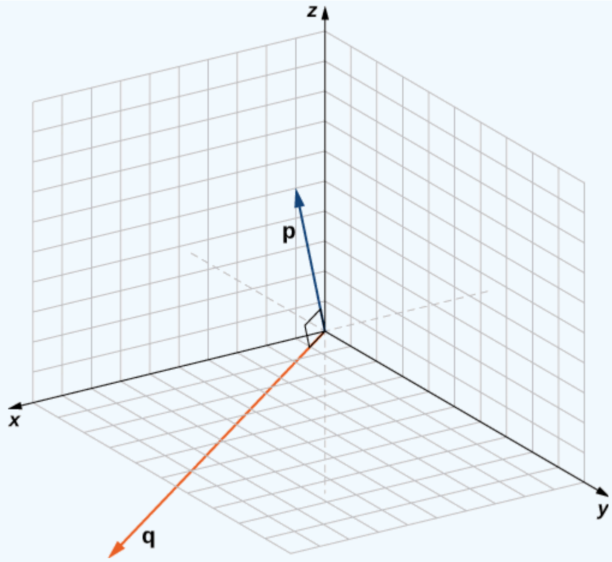


Figure 11.3.4: Vectors $\vec{\mathbf{p}}$ and $\vec{\mathbf{q}}$ form a right angle when their initial points are aligned.

Exercise 11.3.4

For which value of x is $\vec{\mathbf{p}} = \langle 2, 8, -1\rangle$ orthogonal to $\vec{\mathbf{q}} = \langle x, -1, 2\rangle$?

Example 11.3.5: Measuring the Angle Formed by Two Vectors

Let $\vec{\mathbf{v}} = \langle 2, 3, 3\rangle$. Find the measures of the angles formed by the following vectors.

a. $\vec{\mathbf{v}}$ and $\hat{\mathbf{i}}$
b. $\vec{\mathbf{v}}$ and $\hat{\mathbf{j}}$
c. $\vec{\mathbf{v}}$ and $\hat{\mathbf{k}}$

Solution

a.Let α be the angle formed by $\vec{\mathbf{v}}$ and $\hat{\mathbf{i}}$:

$$\begin{aligned}\cos\alpha &= \frac{\vec{\mathbf{v}}\cdot\hat{\mathbf{i}}}{\|\vec{\mathbf{v}}\|\cdot\|\hat{\mathbf{i}}\|} = \frac{\langle 2,3,3\rangle\cdot\langle 1,0,0\rangle}{\sqrt{2^2+3^2+3^2}\sqrt{1}} = \frac{2}{\sqrt{22}} \\ \alpha &= \arccos\frac{2}{\sqrt{22}} \approx 1.130\,\text{rad.}\end{aligned}$$

b. Let β represent the angle formed by $\vec{\mathbf{v}}$ and $\hat{\mathbf{j}}$:

$$\begin{aligned}\cos\beta &= \frac{\vec{\mathbf{v}}\cdot\hat{\mathbf{j}}}{\|\vec{\mathbf{v}}\|\cdot\|\hat{\mathbf{j}}\|} = \frac{\langle 2,3,3\rangle\cdot\langle 0,1,0\rangle}{\sqrt{2^2+3^2+3^2}\sqrt{1}} = \frac{3}{\sqrt{22}} \\ \beta &= \arccos\frac{3}{\sqrt{22}} \approx 0.877\,\text{rad.}\end{aligned}$$

c. Let γ represent the angle formed by $\vec{\mathbf{v}}$ and $\hat{\mathbf{k}}$:

$$\begin{aligned}\cos\gamma &= \frac{\vec{\mathbf{v}}\cdot\hat{\mathbf{k}}}{\|\vec{\mathbf{v}}\|\cdot\|\hat{\mathbf{k}}\|} = \frac{\langle 2,3,3\rangle\cdot\langle 0,0,1\rangle}{\sqrt{2^2+3^2+3^2}\sqrt{1}} = \frac{3}{\sqrt{22}} \\ \gamma &= \arccos\frac{3}{\sqrt{22}} \approx 0.877\,\text{rad.}\end{aligned}$$

Exercise 11.3.5

Let $\vec{\mathbf{v}} = \langle 3, -5, 1\rangle$. Find the measure of the angles formed by each pair of vectors.

a. $\vec{\mathbf{v}}$ and $\hat{\mathbf{i}}$
b. $\vec{\mathbf{v}}$ and $\hat{\mathbf{j}}$
c. $\vec{\mathbf{v}}$ and $\hat{\mathbf{k}}$

The angle a vector makes with each of the coordinate axes, called a direction angle, is very important in practical computations, especially in a field such as engineering. For example, in astronautical engineering, the angle at which a rocket is launched must be determined very precisely. A very small error in the angle can lead to the rocket going hundreds of miles off course. Direction angles are often calculated by using the dot product and the cosines of the angles, called the direction cosines. Therefore, we define both these angles and their cosines.

Definition: direction angles

The angles formed by a nonzero vector and the coordinate axes are called the **direction angles** for the vector (Figure $11.3.5$). The cosines for these angles are called the **direction cosines.**

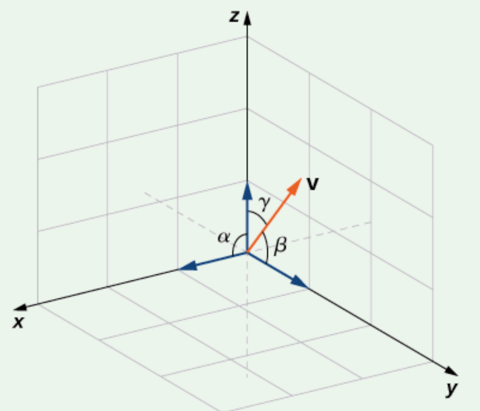


Figure 11.3.5: Angle α is formed by vector $\vec{\mathbf{v}}$ and unit vector $\hat{\mathbf{i}}$. Angle β is formed by vector $\vec{\mathbf{v}}$ and unit vector $\hat{\mathbf{j}}$. Angle γ is formed by vector $\vec{\mathbf{v}}$ and unit vector $\hat{\mathbf{k}}$.

In Example, the direction cosines of $\vec{\mathbf{v}} = \langle 2, 3, 3\rangle$ are $\cos\alpha = \dfrac{2}{\sqrt{22}}, \cos\beta = \dfrac{3}{\sqrt{22}},$ and $\cos\gamma = \dfrac{3}{\sqrt{22}}$. The direction angles of $\vec{\mathbf{v}}$ are $\alpha = 1.130$ rad, $\beta = 0.877$ rad, and $\gamma = 0.877$ rad.

So far, we have focused mainly on vectors related to force, movement, and position in three-dimensional physical space. However, vectors are often used in more abstract ways. For example, suppose a fruit vendor sells apples, bananas, and oranges. On a given day, he sells 30 apples, 12 bananas, and 18 oranges. He might use a quantity vector, $\vec{\mathbf{q}} = \langle 30, 12, 18\rangle$, to represent the quantity of fruit he sold that day. Similarly, he might want to use a price vector, $\vec{\mathbf{p}} = \langle 0.50, 0.25, 1\rangle$,to indicate that he sells his apples for 50¢ each, bananas for 25¢ each, and oranges for $1 apiece. In this example, although we could still graph these vectors, we do not interpret them as literal representations of position in the physical world. We are simply using vectors to keep track of particular pieces of information about apples, bananas, and oranges.

This idea might seem a little strange, but if we simply regard vectors as a way to order and store data, we find they can be quite a powerful tool. Going back to the fruit vendor, let's think about the dot product, $\vec{\mathbf{q}}\cdot\vec{\mathbf{p}}$. We compute it by multiplying the number of apples sold (30) by the price per apple (50¢), the number of bananas sold by the price per banana, and the number of oranges sold by the price per orange. We then add all these values together. So, in this example, the dot product tells us how much money the fruit vendor had in sales on that particular day.

When we use vectors in this more general way, there is no reason to limit the number of components to three. What if the fruit vendor decides to start selling grapefruit? In that case, he would want to use four-dimensional quantity and price vectors to represent the number of apples, bananas, oranges, and grapefruit sold, and their unit prices. As you might expect, to calculate the dot product of four-dimensional vectors, we simply add the products of the components as before, but the sum has four terms instead of three.

Example 11.3.6: Using Vectors in an Economic Context

AAA Party Supply Store sells invitations, party favors, decorations, and food service items such as paper plates and napkins. When AAA buys its inventory, it pays 25¢ per package for invitations and party favors. Decorations cost AAA 50¢ each, and food service items cost 20¢ per package. AAA sells invitations for \$2.50 per package and party favors for \$1.50 per package. Decorations sell for \$4.50 each and food service items for \$1.25 per package.

During the month of May, AAA Party Supply Store sells 1258 invitations, 342 party favors, 2426 decorations, and 1354 food service items. Use vectors and dot products to calculate how much money AAA made in sales during the month of May. How much did the store make in profit?

Solution

The cost, price, and quantity vectors are

$$\begin{aligned}\vec{\mathbf{c}} &= \langle 0.25, 0.25, 0.50, 0.20\rangle \\ \vec{\mathbf{p}} &= \langle 2.50, 1.50, 4.50, 1.25\rangle \\ \vec{\mathbf{q}} &= \langle 1258, 342, 2426, 1354\rangle.\end{aligned}$$

AAA sales for the month of May can be calculated using the dot product $\vec{\mathbf{p}}\cdot\vec{\mathbf{q}}$. We have

$$\begin{aligned}\vec{\mathbf{p}}\cdot\vec{\mathbf{q}} &= \langle 2.50, 1.50, 4.50, 1.25\rangle \cdot \langle 1258, 342, 2426, 1354\rangle \\ &= 3145+513+10917+1692.5 \\ &= 16267.5.\end{aligned}$$

So, AAA took in \$16,267.50 during the month of May. To calculate the profit, we must first calculate how much AAA paid for the items sold. We use the dot product $c\cdot q$ to get

$$\begin{aligned}\vec{\mathbf{c}}\cdot\vec{\mathbf{q}} &= \langle 0.25, 0.25, 0.50, 0.20\rangle \cdot \langle 1258, 342, 2426, 1354\rangle \\ &= 314.5+85.5+1213+270.8 \\ &= 1883.8.\end{aligned}$$

So, AAA paid \$1,883.30 for the items they sold. Their profit, then, is given by

$$\vec{\mathbf{p}}\cdot\vec{\mathbf{q}}-\vec{\mathbf{c}}\cdot\vec{\mathbf{q}} = 16267.5-1883.8 = 14383.7.$$

Therefore, AAA Party Supply Store made \$14,383.70 in May.

Exercise 11.3.6

On June 1, AAA Party Supply Store decided to increase the price they charge for party favors to \$2 per package. They also changed suppliers for their invitations, and are now able to purchase invitations for only 10¢ per package. All their other costs and prices remain the same. If AAA sells 1408 invitations, 147 party favors, 2112 decorations, and 1894 food service items in the month of June, use vectors and dot products to calculate their total sales and profit for June.

Projections

As we have seen, addition combines two vectors to create a resultant vector. But what if we are given a vector and we need to find its component parts? We use vector projections to perform the opposite process; they can break down a vector into its components. The magnitude of a vector projection is a scalar projection. For example, if a child is pulling the handle of a wagon at a 55° angle, we can use projections to determine how much of the force on the handle is actually moving the wagon forward (11.3.6). We return to this example and learn how to solve it after we see how to calculate projections.

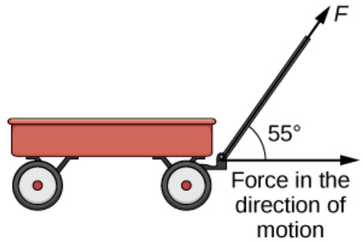


Figure 11.3.6: When a child pulls a wagon, only the horizontal component of the force propels the wagon forward.

Definition: Vector and Projection

The **vector projection** of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$ is the vector labeled $\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}$ in Figure 11.3.7. It has the same initial point as $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ and the same direction as $\vec{\mathbf{u}}$, and represents the component of $\vec{\mathbf{v}}$ that acts in the direction of $\vec{\mathbf{u}}$. If θ represents the angle between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$, then, by properties of triangles, we know the length of $\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}$ is $\|\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}\| = \|\vec{\mathbf{v}}\|\cos\theta$. When expressing $\cos\theta$ in terms of the dot product, this becomes

$$\|\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}\| = \|\vec{\mathbf{v}}\|\cos\theta = \|\vec{\mathbf{v}}\|\left(\frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|}\right) = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|}. \tag{11.3.11}$$

We now multiply by a unit vector in the direction of $\vec{\mathbf{u}}$ to get $\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}$:

$$\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|}\left(\frac{1}{\|\vec{\mathbf{u}}\|}\vec{\mathbf{u}}\right) = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}}. \tag{11.3.12}$$

The length of this vector is also known as the **scalar projection** of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$ and is denoted by

$$\|\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}\| = \text{comp}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|}. \tag{11.3.13}$$

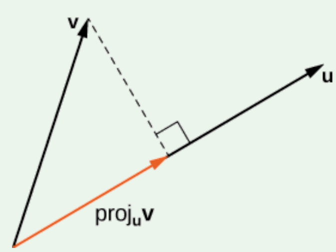


Figure 11.3.7: The projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$ shows the component of vector $\vec{\mathbf{v}}$ in the direction of $\vec{\mathbf{u}}$.

Example 11.3.7: Finding Projections

Find the projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$.

a. $\vec{\mathbf{v}} = \langle 3,5,1\rangle$ and $\vec{\mathbf{u}} = \langle -1,4,3\rangle$
b. $\vec{\mathbf{v}} = 3\hat{\mathbf{i}} - 2\hat{\mathbf{j}}$ and $\vec{\mathbf{u}} = \hat{\mathbf{i}} + 6\hat{\mathbf{j}}$

Solution

a. Substitute the components of $\vec{\mathbf{v}}$ and $\vec{\mathbf{u}}$ into the formula for the projection:

$$\begin{aligned}\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} &= \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}} \\ &= \frac{\langle -1,4,3\rangle\cdot\langle 3,5,1\rangle}{\|\langle -1,4,3\rangle\|^2}\langle -1,4,3\rangle \\ &= \frac{-3+20+3}{(-1)^2+4^2+3^2}\langle -1,4,3\rangle \\ &= \frac{20}{26}\langle -1,4,3\rangle \\ &= \langle -\frac{10}{13},\frac{40}{13},\frac{30}{13}\rangle.\end{aligned}$$

b. To find the two-dimensional projection, simply adapt the formula to the two-dimensional case:

$$\begin{aligned}\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} &= \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}} \\ &= \frac{(\hat{\mathbf{i}}+6\hat{\mathbf{j}})\cdot(3i-2j)}{\left\|\hat{\mathbf{i}}+6\hat{\mathbf{j}}\right\|^2}(i+6j) \\ &= \frac{1(3)+6(-2)}{1^2+6^2}(\hat{\mathbf{i}}+6\hat{\mathbf{j}}) \\ &= -\frac{9}{37}(\hat{\mathbf{i}}+6\hat{\mathbf{j}}) \\ &= -\frac{9}{37}\hat{\mathbf{i}}-\frac{54}{37}\hat{\mathbf{j}}.\end{aligned}$$

Sometimes it is useful to decompose vectors—that is, to break a vector apart into a sum. This process is called the **resolution of a vector into components.** Projections allow us to identify two orthogonal vectors having a desired sum. For example, let $\vec{\mathbf{v}}=\langle 6,-4\rangle$ and let $\vec{\mathbf{u}}=\langle 3,1\rangle$. We want to decompose the vector $\vec{\mathbf{v}}$ into orthogonal components such that one of the component vectors has the same direction as $\vec{\mathbf{u}}$.

We first find the component that has the same direction as $\vec{\mathbf{u}}$ by projecting $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$. Let $\vec{\mathbf{p}}=\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}}$. Then, we have

$$\begin{aligned}\vec{\mathbf{p}} &= \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}} \\ &= \frac{18-4}{9+1}\vec{\mathbf{u}} \\ &= \frac{7}{5}\vec{\mathbf{u}}=\frac{7}{5}\langle 3,1\rangle=\langle\frac{21}{5},\frac{7}{5}\rangle.\end{aligned}$$

Now consider the vector $\vec{\mathbf{q}}=\vec{\mathbf{v}}-\vec{\mathbf{p}}.$ We have

$$\begin{aligned}\vec{\mathbf{q}} &= \vec{\mathbf{v}}-\vec{\mathbf{p}} \\ &= \langle 6,-4\rangle-\langle\frac{21}{5},\frac{7}{5}\rangle \\ &= \langle\frac{9}{5},-\frac{27}{5}\rangle.\end{aligned}$$

Clearly, by the way we defined $\vec{\mathbf{q}}$, we have $\vec{\mathbf{v}}=\vec{\mathbf{q}}+\vec{\mathbf{p}},$ and

$$\begin{aligned}\vec{\mathbf{q}}\cdot\vec{\mathbf{p}} &= \langle\frac{9}{5},-\frac{27}{5}\rangle\cdot\langle\frac{21}{5},\frac{7}{5}\rangle \\ &= \frac{9(21)}{25}+-\frac{27(7)}{25} \\ &= \frac{189}{25}-\frac{189}{25}=0.\end{aligned}$$

Therefore, $\vec{\mathbf{q}}$ and $\vec{\mathbf{p}}$ are orthogonal.

Example 11.3.8: Resolving Vectors into Components

Express $\vec{\mathbf{v}}=\langle 8,-3,-3\rangle$ as a sum of orthogonal vectors such that one of the vectors has the same direction as $\vec{\mathbf{u}}=\langle 2,3,2\rangle$.

Solution

Let $\vec{\mathbf{p}}$ represent the projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$:

$$\begin{aligned}\vec{\mathbf{p}} &= \text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} \\ &= \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}} \\ &= \frac{\langle 2,3,2\rangle\cdot\langle 8,-3,-3\rangle}{\|\langle 2,3,2\rangle\|^2}\langle 2,3,2\rangle \\ &= \frac{16-9-6}{2^2+3^2+2^2}\langle 2,3,2\rangle \\ &= \frac{1}{17}\langle 2,3,2\rangle \\ &= \langle\frac{2}{17},\frac{3}{17},\frac{2}{17}\rangle.\end{aligned}$$

Then,

$$\begin{aligned}\vec{\mathbf{q}} &= \vec{\mathbf{v}}-\vec{\mathbf{p}}=\langle 8,-3,-3\rangle-\langle\frac{2}{17},\frac{3}{17},\frac{2}{17}\rangle \\ &= \langle\frac{134}{17},-\frac{54}{17},-\frac{53}{17}\rangle.\end{aligned}$$

To check our work, we can use the dot product to verify that $\vec{\mathbf{p}}$ and $\vec{\mathbf{q}}$ are orthogonal vectors:

$$\begin{aligned}\vec{\mathbf{p}}\cdot\vec{\mathbf{q}} &= \langle\frac{2}{17},\frac{3}{17},\frac{2}{17}\rangle\cdot\langle\frac{134}{17},-\frac{54}{17},-\frac{53}{17}\rangle \\ &= \frac{268}{17}-\frac{162}{17}-\frac{106}{17}=0.\end{aligned}$$

Then,

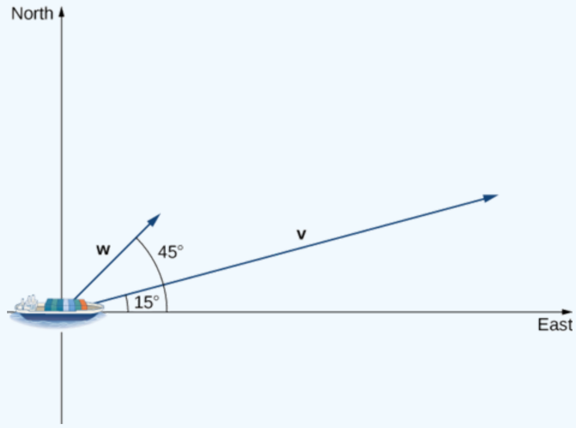
$$\vec{\mathbf{v}}=\vec{\mathbf{p}}+\vec{\mathbf{q}}=\langle\frac{2}{17},\frac{3}{17},\frac{2}{17}\rangle+\langle\frac{134}{17},-\frac{54}{17},-\frac{53}{17}\rangle.$$

Exercise 11.3.7

Express $\vec{\mathbf{v}}=5\,\hat{\mathbf{i}}-\hat{\mathbf{j}}$ as a sum of orthogonal vectors such that one of the vectors has the same direction as $\vec{\mathbf{u}}=4\,\hat{\mathbf{i}}+2\,\hat{\mathbf{j}}$.

Example 11.3.9: Scalar Projection of Velocity

A container ship leaves port traveling 15° north of east. Its engine generates a speed of 20 knots along that path (see the following figure). In addition, the ocean current moves the ship northeast at a speed of 2 knots. Considering both the engine and the current, how fast is the ship moving in the direction 15° north of east? Round the answer to two decimal places.



Solution

Let $\vec{\mathbf{v}}$ be the velocity vector generated by the engine, and let w be the velocity vector of the current. We already know $\|\vec{\mathbf{v}}\| = 20$ along the desired route. We just need to add in the scalar projection of $\vec{\mathbf{w}}$ onto $\vec{\mathbf{v}}$. We get

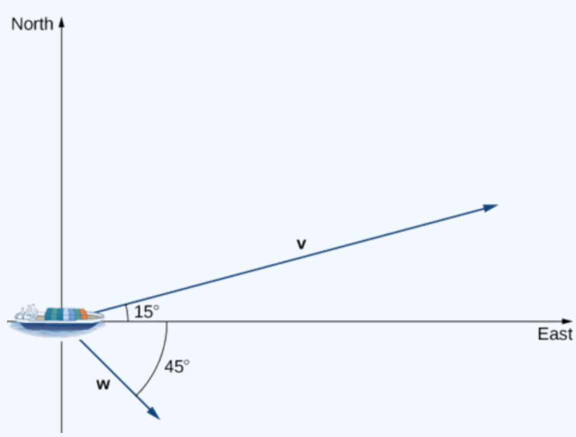
$$\text{comp}_{\vec{\mathbf{v}}}\vec{\mathbf{w}} = \frac{\vec{\mathbf{v}}\cdot\vec{\mathbf{w}}}{\|\vec{\mathbf{v}}\|}$$

$$= \frac{\|\vec{\mathbf{v}}\|\|\vec{\mathbf{w}}\|\cos(30^\circ)}{\|\vec{\mathbf{v}}\|} = \|\vec{\mathbf{w}}\|\cos(30^\circ) = 2\frac{\sqrt{3}}{2} = \sqrt{3} \approx 1.73\,\text{knots}.$$

The ship is moving at 21.73 knots in the direction 15° north of east.

Exercise 11.3.8

Repeat the previous example, but assume the ocean current is moving southeast instead of northeast, as shown in the following figure.



Work

Now that we understand dot products, we can see how to apply them to real-life situations. The most common application of the dot product of two vectors is in the calculation of work.

From physics, we know that work is done when an object is moved by a force. When the force is constant and applied in the same direction the object moves, then we define the work done as the product of the force and the distance the object travels: $W = Fd$. We saw several examples of this type in earlier chapters. Now imagine the direction of the force is different from the direction of motion, as with the example of a child pulling a wagon. To find the work done, we need to multiply the component of the force that acts in the direction of the motion by the magnitude of the displacement. The dot product allows us to do just that. If we represent an applied force by a vector $\vec{\mathbf{F}}$ and the displacement of an object by a vector $\vec{\mathbf{s}}$, then the **work done by the force** is the dot product of $\vec{\mathbf{F}}$ and $\vec{\mathbf{s}}$.

Definition: Constant Force

When a constant force is applied to an object so the object moves in a straight line from point P to point Q, the work W done by the force $\vec{\mathbf{F}}$, acting at an angle θ from the line of motion, is given by

$$W = \vec{\mathbf{F}} \cdot \overrightarrow{PQ} = \| \vec{\mathbf{F}} \| \| \overrightarrow{PQ} \| \cos \theta. \tag{11.3.14}$$

Let's revisit the problem of the child's wagon introduced earlier. Suppose a child is pulling a wagon with a force having a magnitude of 8 lb on the handle at an angle of *55°*. If the child pulls the wagon 50 ft, find the work done by the force (Figure 11.3.8).

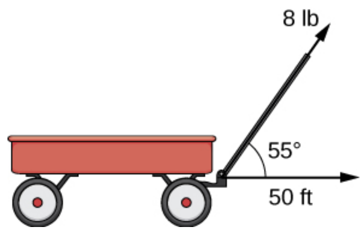


Figure 11.3.8: The horizontal component of the force is the projection of $\vec{\mathbf{F}}$ *onto the positive* x*-axis.*

We have

$$W = \| \vec{\mathbf{F}} \| \| \overrightarrow{PQ} \| \cos \theta = 8(50)(\cos(55^\circ)) \approx 229 \text{ ft}\cdot\text{lb}.$$

In U.S. standard units, we measure the magnitude of force $\left\| \vec{\mathbf{F}} \right\|$ in pounds. The magnitude of the displacement vector $\left\| \overrightarrow{PQ} \right\|$ tells us how far the object moved, and it is measured in feet. The customary unit of measure for work, then, is the foot-pound. One foot-pound is the amount of work required to move an object weighing 1 lb a distance of 1 ft straight up. In the metric system, the unit of measure for force is the newton (N), and the unit of measure of magnitude for work is a newton-meter (N·m), or a joule (J).

Example 11.3.10: Calculating Work

A conveyor belt generates a force $\vec{\mathbf{F}} = 5\,\hat{\mathbf{i}} - 3\,\hat{\mathbf{j}} + \hat{\mathbf{k}}$ that moves a suitcase from point $(1, 1, 1)$ to point $(9, 4, 7)$ along a straight line. Find the work done by the conveyor belt. The distance is measured in meters and the force is measured in newtons.

Solution

The displacement vector $\overrightarrow{PQ}$ has initial point $(1, 1, 1)$ and terminal point $(9, 4, 7)$:

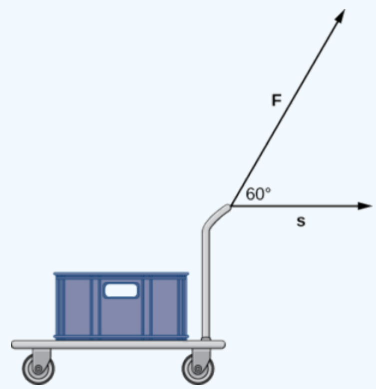
$$\overrightarrow{PQ} = \langle 9-1, 4-1, 7-1 \rangle = \langle 8, 3, 6 \rangle = 8\,\hat{\mathbf{i}} + 3\,\hat{\mathbf{j}} + 6\hat{\mathbf{k}}.$$

Work is the dot product of force and displacement:

$$W = \vec{\mathbf{F}} \cdot \vec{PQ}$$
$$= (5\,\hat{\mathbf{i}} - 3\,\hat{\mathbf{j}} + \hat{\mathbf{k}}) \cdot (8\,\hat{\mathbf{i}} + 3\,\hat{\mathbf{j}} + 6\hat{\mathbf{k}})$$
$$= 5(8) + (-3)(3) + 1(6)$$
$$= 37\,\text{N}\cdot\text{m}$$
$$= 37\,\text{J}$$

Exercise 11.3.9

A constant force of 30 lb is applied at an angle of 60° to pull a handcart 10 ft across the ground. What is the work done by this force?



Key Concepts

- The dot product, or scalar product, of two vectors $\vec{\mathbf{u}} = \langle u_1, u_2, u_3\rangle$ and $\vec{\mathbf{v}} = \langle v_1, v_2, v_3\rangle$ is $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = u_1v_1 + u_2v_2 + u_3v_3$.
- The dot product satisfies the following properties:
  - $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}}$
  - $\vec{\mathbf{u}} \cdot (\vec{\mathbf{v}} + \vec{\mathbf{w}}) = \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} + \vec{\mathbf{u}} \cdot \vec{\mathbf{w}}$
  - $c(\vec{\mathbf{u}} \cdot \vec{\mathbf{v}}) = (c\vec{\mathbf{u}}) \cdot \vec{\mathbf{v}} = \vec{\mathbf{u}} \cdot (c\vec{\mathbf{v}})$
  - $\vec{\mathbf{v}} \cdot \vec{\mathbf{v}} = \|\vec{\mathbf{v}}\|^2$
- The dot product of two vectors can be expressed, alternatively, as $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = \|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|\cos\theta.$ This form of the dot product is useful for finding the measure of the angle formed by two vectors.
- Vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ are orthogonal if $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = 0$.
- The angles formed by a nonzero vector and the coordinate axes are called the *direction angles* for the vector. The cosines of these angles are known as the *direction cosines.*
- The vector projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$ is the vector $\text{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \dfrac{\vec{\mathbf{u}} \cdot \vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}}$. The magnitude of this vector is known as the *scalar projection* of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$, given by $\text{comp}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \dfrac{\vec{\mathbf{u}} \cdot \vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|}$.
- Work is done when a force is applied to an object, causing displacement. When the force is represented by the vector $\vec{\mathbf{F}}$ and the displacement is represented by the vector $\vec{\mathbf{s}}$, then the work done W is given by the formula $W = \vec{\mathbf{F}} \cdot \vec{\mathbf{s}} = \|\vec{\mathbf{F}}\|\,\|\vec{\mathbf{s}}\|\cos\theta.$

Key Equations

- **Dot product of $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$**

  $$\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = u_1v_1 + u_2v_2 + u_3v_3 = \|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|\cos\theta$$

- **Cosine of the angle formed by $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$**

$$\cos\theta = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|\|\vec{\mathbf{v}}\|}$$

- **Vector projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$**

$$\operatorname{proj}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|^2}\vec{\mathbf{u}}$$

- **Scalar projection of $\vec{\mathbf{v}}$ onto $\vec{\mathbf{u}}$**

$$\operatorname{comp}_{\vec{\mathbf{u}}}\vec{\mathbf{v}} = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|}$$

- **Work done by a force $\vec{\mathbf{F}}$ to move an object through displacement vector $\overrightarrow{PQ}$**

$$W = \vec{\mathbf{F}}\cdot\overrightarrow{PQ} = \|\vec{\mathbf{F}}\|\|\overrightarrow{PQ}\|\cos\theta$$

Glossary

Contributors and Attributions

- Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.
- edited for vector notation by Paul Seeburger

Dot Products

Next we learn some vector operations that will be useful to us in doing some geometry. In many ways, vector algebra is the right language for geometry, particularly if we're using functions. In a way, vector algebra is a language and we're using it to express things we've known since childhood. Having a notation for these things will make them more straightforward.

A dot product is a way of multiplying two vectors to get a number, or scalar. Algebraically, suppose $\mathbf{A} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3 \rangle$. We find the *dot product* $\mathbf{A} \cdot \mathbf{B}$ by multiplying the first component of $\mathbf{A}$ by the first component of $\mathbf{B}$, the second component of $\mathbf{A}$ by the second component of $\mathbf{B}$, and so on, and then adding together all these products. So for our sample vectors, $\mathbf{A} \cdot \mathbf{B} = a_1 b_1 + a_2 b_2 + a_3 b_3$. If our vectors have N components, the definition of the dot product becomes:

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^{N} a_i b_i.$$

It is very important to remember that $\mathbf{A} \cdot \mathbf{B}$ is a scalar, not a vector. Also, when writing a dot product we always put a dot symbol between the two vectors to indicate what kind of product we're calculating.

What is it good for? The answer to this question will be clearer after we see a geometric description of the dot product.

Geometrically, the dot product of $\mathbf{A}$ and $\mathbf{B}$ equals the length of $\mathbf{A}$ times the length of $\mathbf{B}$ times the cosine of the angle between them:

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos(\theta).$$

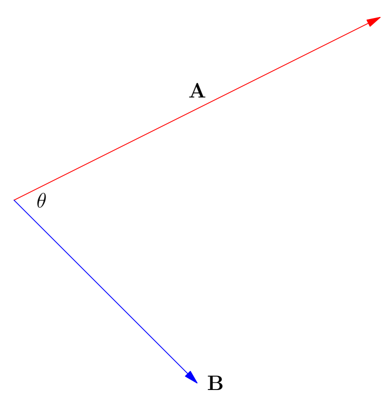


Figure 1: $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos(\theta)$.

This may seem complicated and artificial at first, but we'll find that the dot product gives us useful information about angles and lengths simultaneously. If we've described our vectors using components, the dot product is also easy to calculate.

Equivalence of Descriptions of the Dot Product

We now have two descriptions of the dot product:

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\theta) \qquad \text{and} \qquad \mathbf{A}\cdot\mathbf{B} = \sum_{i=1}^{N} a_i b_i.$$

In mathematics, one tries to justify every statement by proving theorems. The first theorem in 18.02 is to verify that if the dot product is defined to be $\mathbf{A}\cdot\mathbf{B} = \sum_{i=1}^{N} a_i b_i$ then it's also true that $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\theta)$.

We start by multiplying a vector times itself to gain understanding of the geometric definition:

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2\cos(0) = |\mathbf{A}|^2.$$

From the definition of the dot product we get:

$$\mathbf{A}\cdot\mathbf{A} = a_1^2 + a_2^2 + a_3^2 = |\mathbf{A}|^2.$$

The two definitions are equivalent if $\mathbf{A}$ and $\mathbf{B}$ are the same vector.

If $\mathbf{A}$ and $\mathbf{B}$ are different vectors, we can use the law of cosines to show that our geometric description of the dot product of two different vectors is equivalent to its algebraic definition. You may recall that the law of cosines tells you the length of the third side of a triangle given the length of the other two sides and the angle between them.

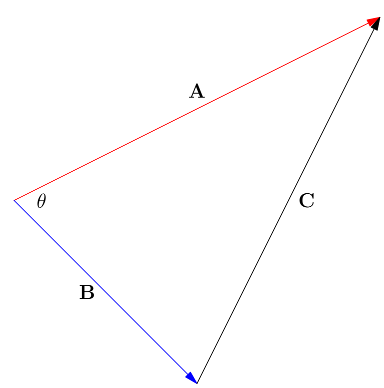


Figure 2: The law of cosines describes the length of $\mathbf{C}$ in terms of $|\mathbf{A}|$, $|\mathbf{B}|$ and θ.

In terms of vectors, the two known sides of our triangle are formed by $\mathbf{A}$ and $\mathbf{B}$. The third side is described by $\mathbf{C} = \mathbf{A} - \mathbf{B}$. The law of cosines then tells us that:

$$|\mathbf{C}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2|\mathbf{A}||\mathbf{B}|\cos\theta.$$

(If you haven't seen this before, then consider what you are about to see to be a proof of the law of cosines based on the assumption that our two descriptions of the dot product are equivalent. If you have seen the law of cosines before, it's the other way around.)

How is the law of cosines related to the dot product?

$$\begin{aligned}
|\mathbf{C}|^2 &= \mathbf{C}\cdot\mathbf{C} \\
&= (\mathbf{A}-\mathbf{B})\cdot(\mathbf{A}-\mathbf{B}) \\
&= \mathbf{A}\cdot\mathbf{A}-\mathbf{A}\cdot\mathbf{B}-\mathbf{B}\cdot\mathbf{A}+\mathbf{B}\cdot\mathbf{B} \\
|\mathbf{C}|^2 &= |\mathbf{A}|^2+|\mathbf{B}|^2-2\mathbf{A}\cdot\mathbf{B}
\end{aligned}$$

Are we allowed to expand $(\mathbf{A}-\mathbf{B})\cdot(\mathbf{A}-\mathbf{B})$ in this way? Is it true that $\mathbf{A}\cdot\mathbf{B}=\mathbf{B}\cdot\mathbf{A}$? Yes; it is possible to prove from the definition of the dot product that commuting, factoring and expanding work with dot products the same way they do with scalar products. (This is where we use the definition of the dot product in this proof.)

Comparing this formula for the length of $\mathbf{C}$ with the one given by the law of cosines, we see that we must have $2\mathbf{A}\cdot\mathbf{B}=2|\mathbf{A}||\mathbf{B}|\cos\theta$, and so we conclude that:

$$\mathbf{A}\cdot\mathbf{B}=|\mathbf{A}||\mathbf{B}|\cos(\theta).$$

Now we have either used the law of cosines to prove that our algebraic and geometric descriptions of the dot product are equivalent, or we have proven the law of cosines based on the assumption that those descriptions are equivalent. A mathematician wold say that the law of cosines is logically equivalent to the statement $\mathbf{A}\cdot\mathbf{B}=|\mathbf{A}||\mathbf{B}|\cos\theta$.

MATH 304

Linear Algebra

Lecture 9:
Properties of determinants.

Determinants

Determinant is a scalar assigned to each square matrix.

Notation. The determinant of a matrix $A = (a_{ij})_{1 \le i,j \le n}$ is denoted $\det A$ or

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Principal property: $\det A \neq 0$ if and only if a system of linear equations with the coefficient matrix A has a unique solution. Equivalently, $\det A \neq 0$ if and only if the matrix A is invertible.

Definition in low dimensions

Definition. $\det(a) = a, \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$- a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

$$+ : \begin{pmatrix} \boxed{*} & * & * \\ * & \boxed{*} & * \\ * & * & \boxed{*} \end{pmatrix}, \begin{pmatrix} * & \boxed{*} & * \\ * & * & \boxed{*} \\ \boxed{*} & * & * \end{pmatrix}, \begin{pmatrix} * & * & \boxed{*} \\ \boxed{*} & * & * \\ * & \boxed{*} & * \end{pmatrix}.$$

$$- : \begin{pmatrix} * & * & \boxed{*} \\ * & \boxed{*} & * \\ \boxed{*} & * & * \end{pmatrix}, \begin{pmatrix} * & \boxed{*} & * \\ \boxed{*} & * & * \\ * & * & \boxed{*} \end{pmatrix}, \begin{pmatrix} \boxed{*} & * & * \\ * & * & \boxed{*} \\ * & \boxed{*} & * \end{pmatrix}.$$

General definition

The general definition of the determinant is quite complicated as there is no simple explicit formula.

There are several approaches to defining determinants.

Approach 1 (original): an explicit (but very complicated) formula.

Approach 2 (axiomatic): we formulate properties that the determinant should have.

Approach 3 (inductive): the determinant of an $n\times n$ matrix is defined in terms of determinants of certain $(n-1)\times(n-1)$ matrices.

$\mathcal{M}_n(\mathbb{R})$: the set of $n \times n$ matrices with real entries.

Theorem There exists a unique function $\det : \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}$ (called the *determinant*) with the following properties:

- if a row of a matrix is multiplied by a scalar r, the determinant is also multiplied by r;
- if we add a row of a matrix multiplied by a scalar to another row, the determinant remains the same;
- if we interchange two rows of a matrix, the determinant changes its sign;
- $\det I = 1$.

Corollary 1 Suppose A is a square matrix and B is obtained from A applying elementary row operations. Then $\det A = 0$ if and only if $\det B = 0$.

Corollary 2 $\det B = 0$ whenever the matrix B has a zero row.

Hint: Multiply the zero row by the zero scalar.

Corollary 3 $\det A = 0$ if and only if the matrix A is not invertible.

Idea of the proof: Let B be the reduced row echelon form of A. If A is invertible then $B = I$; otherwise B has a zero row.

Row echelon form of a square matrix A:

$$\begin{pmatrix} \square & * & * & * & * & * & * \\ & \square & * & * & * & * & * \\ & & \square & * & * & * & * \\ & & & \square & * & * & * \\ & & & & \square & * & * \\ & & & & & \square & * \\ & & & & & & \square \end{pmatrix} \qquad \begin{pmatrix} \square & * & * & * & * & * & * \\ & \square & * & * & * & * & * \\ & & \square & * & * & * & * \\ & & & & \square & * & * \\ & & & & & \square & * \\ & & & & & & \\ & & & & & & \end{pmatrix}$$

$$\det A \neq 0 \qquad\qquad \det A = 0$$

Example. $A = \begin{pmatrix} 3 & -2 & 0 \\ 1 & 0 & 1 \\ -2 & 3 & 0 \end{pmatrix}$, $\det A = ?$

Earlier we have transformed the matrix A into the identity matrix using elementary row operations:

- interchange the 1st row with the 2nd row,
- add -3 times the 1st row to the 2nd row,
- add 2 times the 1st row to the 3rd row,
- multiply the 2nd row by -0.5,
- add -3 times the 2nd row to the 3rd row,
- multiply the 3rd row by -0.4,
- add -1.5 times the 3rd row to the 2nd row,
- add -1 times the 3rd row to the 1st row.

Example. $A = \begin{pmatrix} 3 & -2 & 0 \\ 1 & 0 & 1 \\ -2 & 3 & 0 \end{pmatrix}$, $\det A = ?$

Earlier we have transformed the matrix A into the identity matrix using elementary row operations.

These included two row multiplications, by -0.5 and by -0.4, and one row exchange.

It follows that

$$\det I = -\,(-0.5)\,(-0.4)\det A = (-0.2)\det A.$$

Hence $\det A = -5\det I = -5.$

Other properties of determinants

- If a matrix A has two identical rows then $\det A = 0$.

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \end{vmatrix} = 0$$

- If a matrix A has two rows proportional then $\det A = 0$.

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ ra_1 & ra_2 & ra_3 \end{vmatrix} = r \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \end{vmatrix} = 0$$

Distributive law for rows

- Suppose that matrices X, Y, Z are identical except for the ith row and the ith row of Z is the sum of the ith rows of X and Y.

  Then $\boxed{\det Z = \det X + \det Y.}$

$$\begin{vmatrix} a_1+a_1' & a_2+a_2' & a_3+a_3' \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} + \begin{vmatrix} a_1' & a_2' & a_3' \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

- Adding a scalar multiple of one row to another row does not change the determinant of a matrix.

$$\begin{vmatrix} a_1 + rb_1 & a_2 + rb_2 & a_3 + rb_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} =$$

$$= \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} + \begin{vmatrix} rb_1 & rb_2 & rb_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Definition. A square matrix $A = (a_{ij})$ is called **upper triangular** if all entries below the main diagonal are zeros: $a_{ij} = 0$ whenever $i > j$.

• The determinant of an upper triangular matrix is equal to the product of its diagonal entries.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}$$

• If $A = \mathrm{diag}(d_1, d_2, \dots, d_n)$ then $\det A = d_1 d_2 \dots d_n$. In particular, $\det I = 1$.

Determinant of the transpose

- If A is a square matrix then $\det A^T = \det A$.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Columns vs. rows

- If one column of a matrix is multiplied by a scalar, the determinant is multiplied by the same scalar.

- Interchanging two columns of a matrix changes the sign of its determinant.

- If a matrix A has two columns proportional then $\det A = 0$.

- Adding a scalar multiple of one column to another does not change the determinant of a matrix.

Submatrices

Definition. Given a matrix A, a $k \times k$ **submatrix** of A is a matrix obtained by specifying k columns and k rows of A and deleting the other columns and rows.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 10 & 20 & 30 & 40 \\ 3 & 5 & 7 & 9 \end{pmatrix} \longrightarrow \begin{pmatrix} * & 2 & * & 4 \\ * & * & * & * \\ * & 5 & * & 9 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 4 \\ 5 & 9 \end{pmatrix}$$

Given an $n \times n$ matrix A, let M_{ij} denote the $(n-1)\times(n-1)$ submatrix obtained by deleting the ith row and the jth column of A.

Example. $A = \begin{pmatrix} 3 & -2 & 0 \\ 1 & 0 & 1 \\ -2 & 3 & 0 \end{pmatrix}.$

$$M_{11} = \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix},\ M_{12} = \begin{pmatrix} 1 & 1 \\ -2 & 0 \end{pmatrix},\ M_{13} = \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix},$$

$$M_{21} = \begin{pmatrix} -2 & 0 \\ 3 & 0 \end{pmatrix},\ M_{22} = \begin{pmatrix} 3 & 0 \\ -2 & 0 \end{pmatrix},\ M_{23} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix},$$

$$M_{31} = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix},\ M_{32} = \begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix},\ M_{33} = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}.$$

Row and column expansions

Given an $n \times n$ matrix $A = (a_{ij})$, let M_{ij} denote the $(n-1)\times(n-1)$ submatrix obtained by deleting the ith row and the jth column of A.

Theorem For any $1 \leq k, m \leq n$ we have that

$$\det A = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det M_{kj},$$

$$(\textit{expansion by } k\textit{th row})$$

$$\det A = \sum_{i=1}^{n} (-1)^{i+m} a_{im} \det M_{im}.$$

$$(\textit{expansion by } m\textit{th column})$$

Signs for row/column expansions

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Example. $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

Expansion by the 1st row:

$$\begin{pmatrix} \boxed{1} & * & * \\ * & 5 & 6 \\ * & 8 & 9 \end{pmatrix} \quad \begin{pmatrix} * & \boxed{2} & * \\ 4 & * & 6 \\ 7 & * & 9 \end{pmatrix} \quad \begin{pmatrix} * & * & \boxed{3} \\ 4 & 5 & * \\ 7 & 8 & * \end{pmatrix}$$

$$\det A = 1 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

$$= (5 \cdot 9 - 6 \cdot 8) - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) = 0.$$

Example. $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

Expansion by the 2nd column:

$$\begin{pmatrix} * & \boxed{2} & * \\ 4 & * & 6 \\ 7 & * & 9 \end{pmatrix} \quad \begin{pmatrix} 1 & * & 3 \\ * & \boxed{5} & * \\ 7 & * & 9 \end{pmatrix} \quad \begin{pmatrix} 1 & * & 3 \\ 4 & * & 6 \\ * & \boxed{8} & * \end{pmatrix}$$

$$\det A = -2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 5 \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} - 8 \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}$$

$$= -2(4 \cdot 9 - 6 \cdot 7) + 5(1 \cdot 9 - 3 \cdot 7) - 8(1 \cdot 6 - 3 \cdot 4) = 0.$$

Example. $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

Subtract the 1st row from the 2nd row and from the 3rd row:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 3 & 3 & 3 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 3 & 3 & 3 \\ 6 & 6 & 6 \end{vmatrix} = 0$$

since the last matrix has two proportional rows.

3.3: The Determinant

Learning Objectives

- T/F: The determinant of a matrix is always positive.
- T/F: To compute the determinant of a 3×3 matrix, one needs to compute the determinants of 3 2×2 matrices.
- Give an example of a 2×2 matrix with a determinant of 3.

In this chapter so far we've learned about the transpose (an operation on a matrix that returns another matrix) and the trace (an operation on a square matrix that returns a number). In this section we'll learn another operation on square matrices that returns a number, called the *determinant*. We give a pseudo-definition of the determinant here.

Definition: Determinant

The *determinant* of an $n\times n$ matrix A is a number, denoted $\operatorname{det}(A)$, that is determined by A.

That definition isn't meant to explain everything; it just gets us started by making us realize that the determinant is a number. The determinant is kind of a tricky thing to define. Once you know and understand it, it isn't that hard, but getting started is a bit complicated.[1] We start simply; we define the determinant for 2×2 matrices.

Definition: Determinant of 2×2 Matrices

Let

$$A=\left[\begin{array}{ll}a & b\\ c & d\end{array}\right]. \tag{3.3.1}$$

The *determinant* of A, denoted by

$$\operatorname{det}(A) \text{ or } \left|\begin{array}{ll}a & b\\ c & d\end{array}\right|, \tag{3.3.2}$$

is $ad-bc$.

We've seen the expression $ad-bc$ before. In Section 2.6, we saw that a 2×2 matrix A has inverse

$$\frac{1}{ad-bc}\left[\begin{array}{cc}d & -b\\ -c & a\end{array}\right] \tag{3.3.3}$$

as long as $ad-bc\neq 0$; otherwise, the inverse does not exist. We can rephrase the above statement now: If $\operatorname{det}(A)\neq 0$, then

$$A^{-1}=\frac{1}{\operatorname{det}(A)}\left[\begin{array}{cc}d & -b\\ -c & a\end{array}\right]. \tag{3.3.4}$$

A brief word about the notation: notice that we can refer to the determinant by using what *looks like* absolute value bars around the entries of a matrix. We discussed at the end of the last section the idea of measuring the "size" of a matrix, and mentioned that there are many different ways to measure size. The determinant is one such way. Just as the absolute value of a number measures its size (and ignores its sign), the determinant of a matrix is a measurement of the size of the matrix. (Be careful, though: $\operatorname{det}(A)$ can be negative!)

Let's practice.

Example 3.3.1

Find the determinant of A, B and C where

$$A=\begin{bmatrix}1&2\\3&4\end{bmatrix},\quad B=\begin{bmatrix}3&-1\\2&7\end{bmatrix}\quad\text{and}\quad C=\begin{bmatrix}1&-3\\-2&6\end{bmatrix}. \tag{3.3.5}$$

Solution

Finding the determinant of A:

$$\begin{aligned}\det(A)&=\begin{vmatrix}1&2\\3&4\end{vmatrix}\\&=1(4)-2(3)\\&=-2.\end{aligned} \tag{3.3.6}$$

Similar computations show that $\det(B)=3(7)-(-1)(2)=23$ and $\det(C)=1(6)-(-3)(-2)=0$.

Finding the determinant of a 2×2 matrix is pretty straightforward. It is natural to ask next "How do we compute the determinant of matrices that are not 2×2?" We first need to define some terms.[2]

Definition: Matrix Minor, Cofactor

Let A be an $n\times n$ matrix. The i,j *minor* of A, denoted $A_{i,j}$, is the determinant of the $(n-1)\times(n-1)$ matrix formed by deleting the i^{th} row and j^{th} column of A.

The i,j*-cofactor* of A is the number

$$C_{ij}=(-1)^{i+j}A_{i,j}. \tag{3.3.7}$$

Notice that this definition makes reference to taking the determinant of a matrix, while we haven't yet defined what the determinant is beyond 2×2 matrices. We recognize this problem, and we'll see how far we can go before it becomes an issue.

Examples will help.

Example 3.3.2

Let

$$A=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\quad\text{and}\quad B=\begin{bmatrix}1&2&0&8\\-3&5&7&2\\-1&9&-4&6\\1&1&1&1\end{bmatrix}. \tag{3.3.8}$$

Find $A_{1,3}$, $A_{3,2}$, $B_{2,1}$, $B_{4,3}$ and their respective cofactors.

Solution

To compute the minor $A_{1,3}$, we remove the first row and third column of A then take the determinant.

$$A=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\Rightarrow\begin{bmatrix}\cancel{1}&\cancel{2}&\cancel{3}\\4&5&\cancel{6}\\7&8&\cancel{9}\end{bmatrix}\Rightarrow\begin{bmatrix}4&5\\7&8\end{bmatrix} \tag{3.3.9}$$

$$A_{1,3}=\begin{vmatrix}4&5\\7&8\end{vmatrix}=32-35=-3. \tag{3.3.10}$$

The corresponding cofactor, $C_{1,3}$, is

$$C_{1,3}=(-1)^{1+3}A_{1,3}=(-1)^4(-3)=-3. \tag{3.3.11}$$

The minor $A_{3,2}$ is found by removing the third row and second column of A then taking the determinant.

$$A=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\Rightarrow\begin{bmatrix}1&\cancel{2}&3\\4&\cancel{5}&6\\\cancel{7}&\cancel{8}&\cancel{9}\end{bmatrix}\Rightarrow\begin{bmatrix}1&3\\4&6\end{bmatrix} \tag{3.3.12}$$

$$A_{3,2}=\begin{vmatrix}1&3\\4&6\end{vmatrix}=6-12=-6. \tag{3.3.13}$$

The corresponding cofactor, $C_{3,2}$, is

$$C_{3,2}=(-1)^{3+2}A_{3,2}=(-1)^5(-6)=6. \tag{3.3.14}$$

The minor $B_{2,1}$ is found by removing the second row and first column of B then taking the determinant.

$$B=\begin{bmatrix}1&2&0&8\\-3&5&7&2\\-1&9&-4&6\\1&1&1&1\end{bmatrix}\Rightarrow B=\begin{bmatrix}\cancel{1}&2&0&8\\\cancel{-3}&\cancel{5}&\cancel{7}&\cancel{2}\\\cancel{-1}&9&-4&6\\\cancel{1}&1&1&1\end{bmatrix}\Rightarrow\begin{bmatrix}2&0&8\\9&-4&6\\1&1&1\end{bmatrix} \tag{3.3.15}$$

$$B_{2,1}=\begin{vmatrix}2&0&8\\9&-4&6\\1&1&1\end{vmatrix}\overset{!}{=}? \tag{3.3.16}$$

We're a bit stuck. We don't know how to find the determinate of this 3×3 matrix. We'll come back to this later. The corresponding cofactor is

$$C_{2,1}=(-1)^{2+1}B_{2,1}=-B_{2,1}, \tag{3.3.17}$$

whatever this number happens to be.

The minor $B_{4,3}$ is found by removing the fourth row and third column of B then taking the determinant.

$$B=\begin{bmatrix}1&2&0&8\\-3&5&7&2\\-1&9&-4&6\\1&1&1&1\end{bmatrix}\Rightarrow B=\begin{bmatrix}1&2&\cancel{0}&8\\-3&5&\cancel{7}&2\\-1&9&\cancel{-4}&6\\\cancel{1}&\cancel{1}&\cancel{1}&\cancel{1}\end{bmatrix}\Rightarrow\begin{bmatrix}1&2&8\\-3&5&2\\-1&9&6\end{bmatrix} \tag{3.3.18}$$

$$B_{4,3}=\begin{vmatrix}1&2&8\\-3&5&2\\-1&9&6\end{vmatrix}\overset{!}{=}? \tag{3.3.19}$$

Again, we're stuck. We won't be able to fully compute $C_{4,3}$; all we know so far is that

$$C_{4,3}=(-1)^{4+3}B_{4,3}=(-1)B_{4,3}. \tag{3.3.20}$$

Once we learn how to compute determinates for matrices larger than 2×2 we can come back and finish this exercise.

In our previous example we ran into a bit of trouble. By our definition, in order to compute a minor of an $n\times n$ matrix we needed to compute the determinant of a $(n-1)\times(n-1)$ matrix. This was fine when we started with a 3×3 matrix, but when we got up to a 4×4 matrix (and larger) we run into trouble.

We are almost ready to define the determinant for any square matrix; we need one last definition.

Definition: Cofactor Expansion

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

The *cofactor expansion* of A along the i^{th} *row* is the sum

$$a_{i,1}C_{i,1}+a_{i,2}C_{i,2}+\cdots+a_{i,n}C_{i,n}. \tag{3.3.21}$$

The *cofactor expansion* of A along the j^{th} *row* is the sum

$$a_{1,j}C_{1,j}+a_{2,j}C_{2,j}+\cdots+a_{n,j}C_{n,j}. \tag{3.3.22}$$

The notation of this definition might be a little intimidating, so let's look at an example.

Example 3.3.3

Let

$$A=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}. \tag{3.3.23}$$

Find the cofactor expansions along the second row and down the first column.

Solution

By the definition, the cofactor expansion along the second row is the sum

$$a_{2,1}C_{2,1}+a_{2,2}C_{2,2}+a_{2,3}C_{2,3}. \tag{3.3.24}$$

(Be sure to compare the above line to the definition of cofactor expansion, and see how the "i" in the definition is replaced by "2" here.)

We'll find each cofactor and then compute the sum.

$$C_{2,1}=(-1)^{2+1}\begin{vmatrix}2&3\\8&9\end{vmatrix}=(-1)(-6)=6 \quad \begin{pmatrix}\text{we removed the second row and}\\\text{first column of } A \text{ to compute the}\\\text{minor}\end{pmatrix} \tag{3.3.25}$$

$$C_{2,2}=(-1)^{2+2}\begin{vmatrix}1&3\\7&9\end{vmatrix}=(1)(-12)=-12 \quad \begin{pmatrix}\text{we removed the second row and}\\\text{second column of } A \text{ to compute}\\\text{the minor}\end{pmatrix} \tag{3.3.26}$$

$$C_{2,3}=(-1)^{2+3}\begin{vmatrix}1&2\\7&8\end{vmatrix}=(-1)(-6)=6 \quad \begin{pmatrix}\text{we removed the second row and}\\\text{third column of } A \text{ to compute}\\\text{the minor}\end{pmatrix} \tag{3.3.27}$$

Thus the cofactor expansion along the second row is

$$\begin{aligned}a_{2,1}C_{2,1}+a_{2,2}C_{2,2}+a_{2,3}C_{2,3}&=4(6)+5(-12)+6(6)\\&=24-60+36\\&=0\end{aligned} \tag{3.3.28}$$

At the moment, we don't know what to do with this cofactor expansion; we've just successfully found it.

We move on to find the cofactor expansion down the first column. By the definition, this sum is

$$a_{1,1}C_{1,1}+a_{2,1}C_{2,1}+a_{3,1}C_{3,1}. \tag{3.3.29}$$

(Again, compare this to the above definition and see how we replaced the "j" with "1.")

We find each cofactor:

$$C_{1,1}=(-1)^{1+1}\begin{vmatrix}5&6\\8&9\end{vmatrix}=(1)(-3)=-3 \quad \begin{pmatrix}\text{we removed the first row and first}\\\text{column of } A \text{ to compute the minor}\end{pmatrix} \tag{3.3.30}$$

$$C_{2,1}=(-1)^{2+1}\begin{vmatrix}2&3\\8&9\end{vmatrix}=(-1)(-6)=6 \quad (\text{we computed this cofactor above}) \tag{3.3.31}$$

$$C_{3,1} = (-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = (1)(-3) = -3 \quad \left(\begin{array}{l}\text{we removed the third row and first} \\ \text{column of } A \text{ to compute the minor}\end{array}\right) \tag{3.3.32}$$

The cofactor expansion down the first column is

$$\begin{aligned} a_{1,1}C_{1,1} + a_{2,1}C_{2,1} + a_{3,1}C_{3,1} &= 1(-3) + 4(6) + 7(-3) \\ &= -3 + 24 - 21 \\ &= 0 \end{aligned} \tag{3.3.33}$$

Is it a coincidence that both cofactor expansions were 0? We'll answer that in a while.

This section is entitled "The Determinant," yet we don't know how to compute it yet except for 2×2 matrices. We finally define it now.

Definition: The Determinant

The *determinant* of an $n \times n$ matrix A, denoted $\det(A)$ or $|A|$, is a number given by the following:

- if A is a 1×1 matrix $A = [a]$, then $\det(A) = a$.
- if A is a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{3.3.34}$$

  then $\det(A) = ad - bc$.
- if A is an $n \times n$ matrix, where $n \geq 2$, then $\det(A)$ is the number found by taking the cofactor expansion along the first row of A. That is,

$$\det(A) = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + \cdots + a_{1,n}C_{1,n}. \tag{3.3.35}$$

Notice that in order to compute the determinant of an $n \times n$ matrix, we need to compute the determinants of n $(n-1) \times (n-1)$ matrices. This can be a lot of work. We'll later learn how to shorten some of this. First, let's practice.

Example 3.3.4

Find the determinant of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}. \tag{3.3.36}$$

Solution

Notice that this is the matrix from Example 3.3.3. The cofactor expansion along the first row is

$$\det(A) = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + a_{1,3}C_{1,3}. \tag{3.3.37}$$

We'll compute each cofactor first then take the appropriate sum.

$$\begin{aligned} C_{1,1} &= (-1)^{1+1}A_{1,1} \\ &= 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} \\ &= 45 - 48 \\ &= -3 \end{aligned} \tag{3.3.38}$$

$$\begin{aligned} C_{1,2} &= (-1)^{1+2} A_{1,2} \\ &= (-1) \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} \\ &= (-1)(36-42) \\ &= 6 \end{aligned} \tag{3.3.39}$$

$$\begin{aligned} C_{1,3} &= (-1)^{1+3} A_{1,3} \\ &= 1 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 32-35 \\ &= -3 \end{aligned} \tag{3.3.40}$$

Therefore the determinant of a is

$$\det(A) = 1(-3) + 2(6) + 3(-3) = 0. \tag{3.3.41}$$

Example 3.3.5

Find the determinant of

$$A = \begin{bmatrix} 3 & 6 & 7 \\ 0 & 2 & -1 \\ 3 & -1 & 1 \end{bmatrix}. \tag{3.3.42}$$

Solution

We’ll compute each cofactor first then find the determinant.

$$\begin{aligned} C_{1,1} &= (-1)^{1+1} A_{1,1} \\ &= 1 \cdot \begin{vmatrix} 2 & -1 \\ -1 & 1 \end{vmatrix} \\ &= 2-1 \\ &= 1 \end{aligned} \tag{3.3.43}$$

$$\begin{aligned} C_{1,2} &= (-1)^{1+2} A_{1,2} \\ &= (-1) \cdot \begin{vmatrix} 0 & -1 \\ 3 & 1 \end{vmatrix} \\ &= (-1)(0+3) \\ &= -3 \end{aligned} \tag{3.3.44}$$

$$\begin{aligned} C_{1,3} &= (-1)^{1+3} A_{1,3} \\ &= 1 \cdot \begin{vmatrix} 0 & 2 \\ 3 & -1 \end{vmatrix} \\ &= 0-6 \\ &= -6 \end{aligned} \tag{3.3.45}$$

Thus the determinant is

$$\det(A) = 3(1) + 6(-3) + 7(-6) = -57. \tag{3.3.46}$$

Example 3.3.6

Find the determinant of

$$A=\begin{bmatrix}1 & 2 & 1 & 2\\ -1 & 2 & 3 & 4\\ 8 & 5 & -3 & 1\\ 5 & 9 & -6 & 3\end{bmatrix}. \tag{3.3.47}$$

Solution

This, quite frankly, will take quite a bit of work. In order to compute this determinant, we need to compute 4 minors, each of which requires finding the determinant of a 3×3 matrix! Complaining won't get us any closer to the solution,[3] so let's get started. We first compute the cofactors:

$$\begin{aligned}C_{1,1} &= (-1)^{1+1}A_{1,1}\\ &= 1\cdot\begin{vmatrix}2 & 3 & 4\\ 5 & -3 & 1\\ 9 & -6 & 3\end{vmatrix} \quad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix}\\ &= 2\cdot(-1)^{1+1}\begin{vmatrix}-3 & 1\\ -6 & 3\end{vmatrix}+3\cdot(-1)^{1+2}\begin{vmatrix}5 & 1\\ 9 & 3\end{vmatrix}+4\cdot(-1)^{1+3}\begin{vmatrix}5 & -3\\ 9 & -6\end{vmatrix}\\ &= 2(-3)+3(-6)+4(-3)\\ &= -36\end{aligned} \tag{3.3.48}$$

$$\begin{aligned}C_{1,2} &= (-1)^{1+2}A_{1,2}\\ &= (-1)\cdot\begin{vmatrix}-1 & 3 & 4\\ 8 & -3 & 1\\ 5 & -6 & 3\end{vmatrix} \quad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix}\\ &= (-1)\ \underbrace{\left[(-1)\cdot(-1)^{1+1}\begin{vmatrix}-3 & 1\\ -6 & 3\end{vmatrix}+3\cdot(-1)^{1+2}\begin{vmatrix}8 & 1\\ 5 & 3\end{vmatrix}+4\cdot(-1)^{1+3}\begin{vmatrix}8 & -3\\ 5 & -6\end{vmatrix}\right]}_{\text{the determinant of the } 3\times 3 \text{ matrix}}\\ &= (-1)[(-1)(-3)+3(-19)+4(-33)]\\ &= 186\end{aligned} \tag{3.3.49}$$

$$\begin{aligned}C_{1,3} &= (-1)^{1+3}A_{1,3}\\ &= 1\cdot\begin{vmatrix}-1 & 2 & 4\\ 8 & 5 & 1\\ 5 & 9 & 3\end{vmatrix} \quad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix}\\ &= (-1)\cdot(-1)^{1+1}\begin{vmatrix}5 & 1\\ 9 & 3\end{vmatrix}+2\cdot(-1)^{1+2}\begin{vmatrix}8 & 1\\ 5 & 3\end{vmatrix}+4\cdot(-1)^{1+3}\begin{vmatrix}8 & 5\\ 5 & 9\end{vmatrix}\\ &= (-1)(6)+2(-19)+4(47)\\ &= 144\end{aligned} \tag{3.3.50}$$

$$\begin{aligned} C_{1,4} &= (-1)^{1+4}A_{1,4} \\ &= (-1)\cdot\begin{vmatrix} -1 & 2 & 3 \\ 8 & 5 & -3 \\ 5 & 9 & -6 \end{vmatrix} \quad \left(\begin{array}{c}\text{we must compute the determinant} \\ \text{of this } 3\times 3 \text{ matrix}\end{array}\right) \\ &= (-1)\ \underbrace{\left[(-1)\cdot(-1)^{1+1}\begin{vmatrix} 5 & -3 \\ 9 & -6 \end{vmatrix} + 2\cdot(-1)^{1+2}\begin{vmatrix} 8 & -3 \\ 5 & -6 \end{vmatrix} + 3\cdot(-1)^{1+3}\begin{vmatrix} 8 & 5 \\ 5 & 9 \end{vmatrix}\right]}_{\text{the determinant of the } 3\times 3 \text{ matrix}} \\ &= (-1)[(-1)(-3)+2(33)+3(47)] \\ &= -210 \end{aligned} \tag{3.3.51}$$

We’ve computed our four cofactors. All that is left is to compute the cofactor expansion.

$$\det(A) = 1(-36)+2(186)+1(144)+2(-210) = 60. \tag{3.3.52}$$

As a way of “visualizing” this, let’s write out the cofactor expansion again but including the matrices in their place.

$$\begin{aligned} \det(A) &= a_{1,1}C_{1,1}+a_{1,2}C_{1,2}+a_{1,3}C_{1,3}+a_{1,4}C_{1,4} \\ &= 1(-1)^2\underbrace{\begin{vmatrix} 2 & 3 & 4 \\ 5 & -3 & 1 \\ 9 & -6 & 3 \end{vmatrix}}_{=-36} + 2(-1)^3\underbrace{\begin{vmatrix} -1 & 3 & 4 \\ 8 & -3 & 1 \\ 5 & -6 & 3 \end{vmatrix}}_{=-186} \\ &\quad + 1(-1)^4\underbrace{\begin{vmatrix} -1 & 2 & 4 \\ 8 & 5 & 1 \\ 5 & 9 & 3 \end{vmatrix}}_{=144} + 2(-1)^5\underbrace{\begin{vmatrix} -1 & 2 & 3 \\ 8 & 5 & -3 \\ 5 & 9 & -6 \end{vmatrix}}_{=210} \\ &= 60 \end{aligned} \tag{3.3.53}$$

That certainly took a while; it required more than 50 multiplications (we didn’t count the additions). To compute the determinant of a 5×5 matrix, we’ll need to compute the determinants of five 4×4 matrices, meaning that we’ll need over 250 multiplications! Not only is this a lot of work, but there are just too many ways to make silly mistakes.[4] There are some tricks to make this job easier, but regardless we see the need to employ technology. Even then, technology quickly bogs down. A 25×25 matrix is considered “small” by today’s standards,[5] but it is essentially impossible for a computer to compute its determinant by only using cofactor expansion; it too needs to employ “tricks.”

In the next section we will learn some of these tricks as we learn some of the properties of the determinant. Right now, let’s review the essentials of what we have learned.

1. The determinant of a square matrix is a number that is determined by the matrix.
2. We find the determinant by computing the cofactor expansion along the first row.
3. To compute the determinant of an $n\times n$ matrix, we need to compute n determinants of $(n-1)\times(n-1)$ matrices.

Footnotes

[1] It’s similar to learning to ride a bike. The riding itself isn’t hard, it is getting started that’s difficult.

[2] This is the standard definition of these two terms, although slight variations exist.

[3] But it might make us feel a little better. Glance ahead: do you see how much work we have to do?!?

[4] The author made three when the above example was originally typed.

[5] It is common for mathematicians, scientists and engineers to consider linear systems with thousands of equations and variables.

Properties of determinants

Determinants

Now halfway through the course, we leave behind rectangular matrices and focus on square ones. Our next big topics are determinants and eigenvalues.

The *determinant* is a number associated with any square matrix; we'll write it as $\det A$ or $|A|$. The determinant encodes a lot of information about the matrix; the matrix is invertible exactly when the determinant is non-zero.

Properties

Rather than start with a big formula, we'll list the properties of the determinant. We already know that $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$; these properties will give us a formula for the determinant of square matrices of all sizes.

1. $\det I = 1$
2. If you exchange two rows of a matrix, you reverse the sign of its determinant from positive to negative or from negative to positive.
3. (a) If we multiply one row of a matrix by t, the determinant is multiplied by t: $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

   (b) The determinant behaves like a linear function on the rows of the matrix:
$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

Property 1 tells us that $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$. Property 2 tells us that $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$. The determinant of a permutation matrix P is 1 or -1 depending on whether P exchanges an even or odd number of rows.

From these three properties we can deduce many others:

4. If two rows of a matrix are equal, its determinant is zero.

   This is because of property 2, the exchange rule. On the one hand, exchanging the two identical rows does not change the determinant. On the other hand, exchanging the two rows changes the sign of the determinant. Therefore the determinant must be 0.

5. If $i \neq j$, subtracting t times row i from row j doesn't change the determinant.

In two dimensions, this argument looks like:

$$\begin{aligned}
\begin{vmatrix} a & b \\ c-ta & d-tb \end{vmatrix} &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \begin{vmatrix} a & b \\ ta & tb \end{vmatrix} && \text{property 3(b)} \\
&= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - t\begin{vmatrix} a & b \\ a & b \end{vmatrix} && \text{property 3(a)} \\
&= \begin{vmatrix} a & b \\ c & d \end{vmatrix} && \text{property 4.}
\end{aligned}$$

The proof for higher dimensional matrices is similar.

6. If A has a row that is all zeros, then $\det A = 0$.

   We get this from property 3 (a) by letting $t = 0$.

7. The determinant of a triangular matrix is the product of the diagonal entries (pivots) $d_1, d_2, ..., d_n$.

   Property 5 tells us that the determinant of the triangular matrix won't change if we use elimination to convert it to a diagonal matrix with the entries d_i on its diagonal. Then property 3 (a) tells us that the determinant of this diagonal matrix is the product $d_1 d_2 \cdots d_n$ times the determinant of the identity matrix. Property 1 completes the argument.

   Note that we cannot use elimination to get a diagonal matrix if one of the d_i is zero. In that case elimination will give us a row of zeros and property 6 gives us the conclusion we want.

8. $\det A = 0$ exactly when A is singular.

   If A is singular, then we can use elimination to get a row of zeros, and property 6 tells us that the determinant is zero.

   If A is not singular, then elimination produces a full set of pivots $d_1, d_2, ..., d_n$ and the determinant is $d_1 d_2 \cdots d_n \neq 0$ (with minus signs from row exchanges).

We now have a very practical formula for the determinant of a non-singular matrix. In fact, the way computers find the determinants of large matrices is to first perform elimination (keeping track of whether the number of row exchanges is odd or even) and then multiply the pivots:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{bmatrix}, \text{ if } a \neq 0, \text{ so}$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a(d - \frac{c}{a}b) = ad - bc.$$

9. $\det AB = (\det A)(\det B)$

   This is very useful. Although the determinant of a sum does not equal the sum of the determinants, it is true that the determinant of a product equals the product of the determinants.

For example:

$$\det A^{-1} = \frac{1}{\det A},$$

because $A^{-1}A = 1$. (Note that if A is singular then A^{-1} does not exist and $\det A^{-1}$ is undefined.) Also, $\det A^2 = (\det A)^2$ and $\det 2A = 2^n \det A$ (applying property 3 to each row of the matrix). This reminds us of volume – if we double the length, width and height of a three dimensional box, we increase its volume by a multiple of $2^3 = 8$.

10. $\det A^T = \det A$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc.$$

This lets us translate properties (2, 3, 4, 5, 6) involving rows into statements about columns. For instance, if a column of a matrix is all zeros then the determinant of that matrix is zero.

To see why $|A^T| = |A|$, use elimination to write $A = LU$. The statement becomes $|U^T L^T| = |LU|$. Rule 9 then tells us $|U^T||L^T| = |L||U|$.

Matrix L is a lower triangular matrix with 1's on the diagonal, so rule 5 tells us that $|L| = |L^T| = 1$. Because U is upper triangular, rule 5 tells us that $|U| = |U^T|$. Therefore $|U^T||L^T| = |L||U|$ and $|A^T| = |A|$.

We have one loose end to worry about. Rule 2 told us that a row exchange changes the sign of the determinant. If it's possible to do seven row exchanges and get the same matrix you would by doing ten row exchanges, then we could prove that the determinant equals its negative. To complete the proof that the determinant is well defined by properties 1, 2 and 3 we'd need to show that the result of an odd number of row exchanges (odd permutation) can never be the same as the result of an even number of row exchanges (even permutation).

2.5 Inverse Matrices

Suppose A is a square matrix. We look for an "***inverse matrix***" A^{-1} of the same size, such that A^{-1} *times* A *equals* I. Whatever A does, A^{-1} undoes. Their product is the identity matrix—which does nothing to a vector, so $A^{-1}A\boldsymbol{x} = \boldsymbol{x}$. *But* A^{-1} *might not exist.*

What a matrix mostly does is to multiply a vector $\boldsymbol{x}$. Multiplying $A\boldsymbol{x} = \boldsymbol{b}$ by A^{-1} gives $A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}$. ***This is*** $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. The product $A^{-1}A$ is like multiplying by a number and then dividing by that number. A number has an inverse if it is not zero—matrices are more complicated and more interesting. The matrix A^{-1} is called "A inverse."

DEFINITION The matrix A is ***invertible*** if there exists a matrix A^{-1} such that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I. \tag{1}$$

Not all matrices have inverses. This is the first question we ask about a square matrix: Is A invertible? We don't mean that we immediately calculate A^{-1}. In most problems we never compute it! Here are six "notes" about A^{-1}.

Note 1 ***The inverse exists if and only if elimination produces*** n ***pivots*** (row exchanges are allowed). Elimination solves $A\boldsymbol{x} = \boldsymbol{b}$ without explicitly using the matrix A^{-1}.

Note 2 The matrix A cannot have two different inverses. Suppose $BA = I$ and also $AC = I$. Then $B = C$, according to this "proof by parentheses":

$$B(AC) = (BA)C \quad \text{gives} \quad BI = IC \quad \text{or} \quad B = C. \tag{2}$$

This shows that a *left-inverse* B (multiplying from the left) and a *right-inverse* C (multiplying A from the right to give $AC = I$) must be the *same matrix.*

Note 3 If A is invertible, the one and only solution to $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{x} = A^{-1}\boldsymbol{b}$:

$$\textbf{\textit{Multiply}} \quad A\boldsymbol{x} = \boldsymbol{b} \quad \textbf{\textit{by}} \quad A^{-1}. \quad \textbf{\textit{Then}} \quad \boldsymbol{x} = A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}.$$

Note 4 (Important) ***Suppose there is a nonzero vector*** $\boldsymbol{x}$ ***such that*** $A\boldsymbol{x} = \boldsymbol{0}$. *Then* A *cannot have an inverse.* No matrix can bring $\boldsymbol{0}$ back to $\boldsymbol{x}$.

If A is invertible, then $A\boldsymbol{x} = \boldsymbol{0}$ can only have the zero solution $\boldsymbol{x} = A^{-1}\boldsymbol{0} = \boldsymbol{0}$.

Note 5 A 2 by 2 matrix is invertible if and only if $ad - bc$ is not zero:

$$\textbf{2 by 2 Inverse:} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{3}$$

This number $ad - bc$ is the *determinant* of A. A matrix is invertible if its determinant is not zero (Chapter 5). The test for n pivots is usually decided before the determinant appears.

Note 6 A diagonal matrix has an inverse provided no diagonal entries are zero:

$$\text{If} \quad A = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \quad \text{then} \quad A^{-1} = \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{bmatrix}.$$

Example 1 The 2 by 2 matrix $A = \left[\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix}\right]$ is not invertible. It fails the test in Note 5, because $ad - bc$ equals $2 - 2 = 0$. It fails the test in Note 3, because $A\boldsymbol{x} = \mathbf{0}$ when $\boldsymbol{x} = (2, -1)$. It fails to have two pivots as required by Note 1.

Elimination turns the second row of this matrix A into a zero row.

The Inverse of a Product AB

For two nonzero numbers a and b, the sum $a + b$ might or might not be invertible. The numbers $a = 3$ and $b = -3$ have inverses $\frac{1}{3}$ and $-\frac{1}{3}$. Their sum $a + b = 0$ has no inverse. But the product $ab = -9$ does have an inverse, which is $\frac{1}{3}$ times $-\frac{1}{3}$.

For two matrices A and B, the situation is similar. It is hard to say much about the invertibility of $A + B$. But the *product* AB has an inverse, if and only if the two factors A and B are separately invertible (and the same size). The important point is that A^{-1} and B^{-1} come in ***reverse order***:

If A and B are invertible then so is AB. The inverse of a product AB is

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{4}$$

To see why the order is reversed, multiply AB times $B^{-1}A^{-1}$. Inside that is $BB^{-1} = I$:

$$\textbf{Inverse of } AB \qquad (AB)(B^{-1}A^{-1}) = AIA^{-1} = AA^{-1} = I.$$

We moved parentheses to multiply BB^{-1} first. Similarly $B^{-1}A^{-1}$ times AB equals I. This illustrates a basic rule of mathematics: Inverses come in reverse order. It is also common sense: If you put on socks and then shoes, the first to be taken off are the _____. The same reverse order applies to three or more matrices:

$$\textbf{Reverse order} \qquad (ABC)^{-1} = C^{-1}B^{-1}A^{-1}. \tag{5}$$

Example 2 ***Inverse of an elimination matrix***. If E subtracts 5 times row 1 from row 2, then E^{-1} *adds* 5 times row 1 to row 2:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Multiply EE^{-1} to get the identity matrix I. Also multiply $E^{-1}E$ to get I. We are adding and subtracting the same 5 times row 1. Whether we add and then subtract (this is EE^{-1}) or subtract and then add (this is $E^{-1}E$), we are back at the start.

For square matrices, an inverse on one side is automatically an inverse on the other side. If $AB = I$ then automatically $BA = I$. In that case B is A^{-1}. This is very useful to know but we are not ready to prove it.

Example 3 Suppose F subtracts 4 times row 2 from row 3, and F^{-1} adds it back:

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \quad \text{and} \quad F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}.$$

Now multiply F by the matrix E in Example 2 to find FE. Also multiply E^{-1} times F^{-1} to find $(FE)^{-1}$. Notice the orders FE and $E^{-1}F^{-1}$!

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 20 & -4 & 1 \end{bmatrix} \quad \text{is inverted by} \quad E^{-1}F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{5} & 1 & 0 \\ 0 & \mathbf{4} & 1 \end{bmatrix}. \tag{6}$$

The result is beautiful and correct. The product FE contains "20" but its inverse doesn't. E subtracts 5 times row 1 from row 2. Then F subtracts 4 times the *new* row 2 (changed by row 1) from row 3. ***In this order*** FE***, row 3 feels an effect from row 1.***

In the order $E^{-1}F^{-1}$, that effect does not happen. First F^{-1} adds 4 times row 2 to row 3. After that, E^{-1} adds 5 times row 1 to row 2. There is no 20, because row 3 doesn't change again. ***In this order*** $E^{-1}F^{-1}$***, row*** 3 ***feels no effect from row 1.***

In elimination order F follows E. In reverse order E^{-1} follows F^{-1}.
$\boldsymbol{E^{-1}F^{-1}}$ ***is quick. The multipliers 5, 4 fall into place below the diagonal of*** $\mathbf{1}$***'s.***

This special multiplication $E^{-1}F^{-1}$ and $E^{-1}F^{-1}G^{-1}$ will be useful in the next section. We will explain it again, more completely. In this section our job is A^{-1}, and we expect some serious work to compute it. Here is a way to organize that computation.

Calculating A^{-1} by Gauss-Jordan Elimination

I hinted that A^{-1} might not be explicitly needed. The equation $A\boldsymbol{x} = \boldsymbol{b}$ is solved by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. But it is not necessary or efficient to compute A^{-1} and multiply it times $\boldsymbol{b}$. *Elimination goes directly to* $\boldsymbol{x}$. Elimination is also the way to calculate A^{-1}, as we now show. The Gauss-Jordan idea is to solve $AA^{-1} = I$, *finding each column of* A^{-1}.

A multiplies the first column of A^{-1} (call that $\boldsymbol{x}_1$) to give the first column of I (call that $\boldsymbol{e}_1$). This is our equation $A\boldsymbol{x}_1 = \boldsymbol{e}_1 = (1, 0, 0)$. There will be two more equations. Each of the columns $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ of A^{-1} is multiplied by A to produce a column of I:

$$\textbf{3 columns of } A^{-1} \qquad AA^{-1} = A\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{bmatrix} = I. \tag{7}$$

To invert a 3 by 3 matrix A, we have to solve three systems of equations: $A\boldsymbol{x}_1 = \boldsymbol{e}_1$ and $A\boldsymbol{x}_2 = \boldsymbol{e}_2 = (0, 1, 0)$ and $A\boldsymbol{x}_3 = \boldsymbol{e}_3 = (0, 0, 1)$. Gauss-Jordan finds A^{-1} this way.

The ***Gauss-Jordan method*** computes A^{-1} by solving ***all n equations together***. Usually the "augmented matrix" $[A \;\; \boldsymbol{b}]$ has one extra column $\boldsymbol{b}$. Now we have three right sides $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ (when A is 3 by 3). They are the columns of I, so the augmented matrix is really the block matrix $[\,A \;\; I\,]$. I take this chance to invert my favorite matrix K, with 2's on the main diagonal and -1's next to the 2's:

$$\begin{bmatrix} K & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad \textbf{Start Gauss-Jordan on } K$$

$$\rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{-1} & \frac{1}{2} & \mathbf{1} & \mathbf{0} \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad (\tfrac{1}{2}\;\textbf{\textit{row}}\;\mathbf{1} + \textbf{\textit{row}}\;\mathbf{2})$$

$$\rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & \mathbf{1} \end{bmatrix} \quad (\tfrac{2}{3}\;\textbf{\textit{row}}\;\mathbf{2} + \textbf{\textit{row}}\;\mathbf{3})$$

We are halfway to K^{-1}. The matrix in the first three columns is U (upper triangular). The pivots $2, \frac{3}{2}, \frac{4}{3}$ are on its diagonal. Gauss would finish by back substitution. The contribution of Jordan is *to continue with elimination*! He goes all the way to the ***"reduced echelon form"***. Rows are added to rows above them, to produce ***zeros above the pivots***:

$$\begin{pmatrix} \text{Zero above} \\ \text{third pivot} \end{pmatrix} \rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{0} & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{3}{4}\;\textbf{\textit{row}}\;\mathbf{3} + \textbf{\textit{row}}\;\mathbf{2})$$

$$\begin{pmatrix} \text{Zero above} \\ \text{second pivot} \end{pmatrix} \rightarrow \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} & \frac{3}{2} & \mathbf{1} & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{2}{3}\;\textbf{\textit{row}}\;\mathbf{2} + \textbf{\textit{row}}\;\mathbf{1})$$

The last Gauss-Jordan step is to divide each row by its pivot. The new pivots are 1. We have reached I in the first half of the matrix, because K is invertible. ***The three columns of K^{-1} are in the second half of $[\,I \;\; K^{-1}\,]$***:

$$\begin{matrix} (\text{divide by } 2) \\ (\text{divide by } \frac{3}{2}) \\ (\text{divide by } \frac{4}{3}) \end{matrix} \quad \begin{bmatrix} \mathbf{1} & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \mathbf{1} & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \mathbf{1} & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} I & \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} I & \boldsymbol{K^{-1}} \end{bmatrix}.$$

Starting from the 3 by 6 matrix $[\,K \;\; I\,]$, we ended with $[\,I \;\; K^{-1}\,]$. Here is the whole Gauss-Jordan process on one line for any invertible matrix A:

Gauss-Jordan ***Multiply*** $\boldsymbol{[\,A \;\; I\,]}$ ***by*** $\boldsymbol{A^{-1}}$ ***to get*** $\boldsymbol{[\,I \;\; A^{-1}\,]}$.

The elimination steps create the inverse matrix while changing A to I. For large matrices, we probably don't want A^{-1} at all. But for small matrices, it can be very worthwhile to know the inverse. We add three observations about this particular K^{-1} because it is an important example. We introduce the words *symmetric*, *tridiagonal*, and *determinant*:

1. K is ***symmetric*** across its main diagonal. So is K^{-1}.

2. K is ***tridiagonal*** (only three nonzero diagonals). But K^{-1} is a dense matrix with no zeros. That is another reason we don't often compute inverse matrices. The inverse of a band matrix is generally a dense matrix.

3. The *product of pivots* is $2(\frac{3}{2})(\frac{4}{3}) = 4$. This number 4 is the ***determinant*** of K.

K^{-1} involves division by the determinant
$$K^{-1} = \frac{1}{4}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}. \tag{8}$$

This is why an invertible matrix cannot have a zero determnant.

Example 4 Find A^{-1} by Gauss-Jordan elimination starting from $A = \left[\begin{smallmatrix} 2 & 3 \\ 4 & 7 \end{smallmatrix}\right]$. There are two row operations and then a division to put 1's in the pivots:

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{3} & 1 & 0 \\ \mathbf{4} & \mathbf{7} & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 2 & 3 & 1 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{1} \end{bmatrix} \quad (\text{this is } \begin{bmatrix} U & L^{-1} \end{bmatrix})$$
$$\to \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{7} & \mathbf{-3} \\ 0 & 1 & -2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{7}{2} & -\frac{3}{2} \\ 0 & 1 & -2 & 1 \end{bmatrix} \quad (\text{this is } \begin{bmatrix} I & \boldsymbol{A^{-1}} \end{bmatrix}).$$

That A^{-1} involves division by the determinant $ad - bc = 2 \cdot 7 - 3 \cdot 4 = 2$. The code for $X =$ inverse(A) can use **rref**, the "row reduced echelon form" from Chapter 3:

$I =$ eye (n);	% Define the n by n identity matrix
$R =$ rref $([A\ I])$;	% Eliminate on the augmented matrix $[A\ I]$
$X = R(:, n+1 : n+n)$	% Pick A^{-1} from the last n columns of R

A must be invertible, or elimination cannot reduce it to I (in the left half of R).

Gauss-Jordan shows why A^{-1} is expensive. We must solve n equations for its n columns.

To solve $Ax = b$ without A^{-1}, we deal with *one* column b to find one column x.

In defense of A^{-1}, we want to say that its cost is not n times the cost of one system $A\boldsymbol{x} = \boldsymbol{b}$. Surprisingly, the cost for n columns is only multiplied by 3. This saving is because the n equations $A\boldsymbol{x}_i = \boldsymbol{e}_i$ all involve the same matrix A. Working with the right sides is relatively cheap, because elimination only has to be done once on A.

The complete A^{-1} needs n^3 elimination steps, where a single $\boldsymbol{x}$ needs $n^3/3$. The next section calculates these costs.

Singular versus Invertible

We come back to the central question. Which matrices have inverses? The start of this section proposed the pivot test: ***A^{-1} exists exactly when A has a full set of n pivots.*** (Row exchanges are allowed.) Now we can prove that by Gauss-Jordan elimination:

1. With n pivots, elimination solves all the equations $A\boldsymbol{x}_i = \boldsymbol{e}_i$. The columns $\boldsymbol{x}_i$ go into A^{-1}. Then $AA^{-1} = I$ and A^{-1} is at least a ***right-inverse***.

2. Elimination is really a sequence of multiplications by E's and P's and D^{-1}:

Left-inverse $$(D^{-1} \cdots E \cdots P \cdots E)A = I. \tag{9}$$

D^{-1} divides by the pivots. The matrices E produce zeros below and above the pivots. P will exchange rows if needed (see Section 2.7). The product matrix in equation (9) is evidently a ***left-inverse***. With n pivots we have reached $A^{-1}A = I$.

The right-inverse equals the left-inverse. That was Note 2 at the start of in this section. So a square matrix with a full set of pivots will always have a two-sided inverse.

Reasoning in reverse will now show that A ***must have n pivots if*** $AC = I$. (Then we deduce that C is also a left-inverse and $CA = I$.) Here is one route to those conclusions:

1. If A doesn't have n pivots, elimination will lead to a *zero row*.
2. Those elimination steps are taken by an invertible M. *So a row of MA is zero.*
3. If $AC = I$ had been possible, then $MAC = M$. The zero row of MA, times C, gives a zero row of M itself.
4. An invertible matrix M can't have a zero row! A *must* have n pivots if $AC = I$.

That argument took four steps, but the outcome is short and important.

Elimination gives a complete test for invertibility of a square matrix. ***A^{-1} exists (and Gauss-Jordan finds it) exactly when A has n pivots***. The argument above shows more:

$$\text{If} \quad AC = I \quad \text{then} \quad CA = I \quad \text{and} \quad C = A^{-1}$$

Example 5 If L is lower triangular with 1's on the diagonal, so is L^{-1}.

A triangular matrix is invertible if and only if no diagonal entries are zero.

Here L has 1's so L^{-1} also has 1's. Use the Gauss-Jordan method to construct L^{-1}. Start by subtracting multiples of pivot rows from rows *below*. Normally this gets us halfway to the inverse, but for L it gets us all the way. L^{-1} appears on the right when I appears on the left. Notice how L^{-1} contains 11, from 3 times 5 minus 4.

Gauss-Jordan on triangular L

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 0 & 0 \\ \mathbf{3} & \mathbf{1} & \mathbf{0} & 0 & 1 & 0 \\ \mathbf{4} & \mathbf{5} & \mathbf{1} & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} L & I \end{bmatrix}$$

$$\begin{matrix} \\ \to \\ \to \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 5 & 1 & -4 & 0 & 1 \end{bmatrix} \quad \begin{matrix} \text{(3 times row 1 from row 2)} \\ \text{(4 times row 1 from row 3)} \\ \text{(then 5 times row 2 from row 3)} \end{matrix}$$

$$\begin{matrix} \\ \\ \to \end{matrix} \begin{bmatrix} 1 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{-3} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{11} & \mathbf{-5} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} I & \boldsymbol{L^{-1}} \end{bmatrix}.$$

L goes to I by a product of elimination matrices $E_{32}E_{31}E_{21}$. So that product is L^{-1}. All pivots are 1's (a full set). L^{-1} is lower triangular, with the strange entry "11".

That 11 does not appear to spoil 3, 4, 5 in the good order $E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = L$.

■ REVIEW OF THE KEY IDEAS ■

1. The inverse matrix gives $AA^{-1} = I$ and $A^{-1}A = I$.
2. A is invertible if and only if it has n pivots (row exchanges allowed).
3. If $A\boldsymbol{x} = \mathbf{0}$ for a nonzero vector $\boldsymbol{x}$, then A has no inverse.
4. The inverse of AB is the reverse product $B^{-1}A^{-1}$. And $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.
5. The Gauss-Jordan method solves $AA^{-1} = I$ to find the n columns of A^{-1}. The augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$ is row-reduced to $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

■ WORKED EXAMPLES ■

2.5 A The inverse of a triangular **difference matrix** A is a triangular **sum matrix** S:

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] = \begin{bmatrix} I & A^{-1} \end{bmatrix} = \begin{bmatrix} I & \textit{sum matrix} \end{bmatrix}.$$

If I change a_{13} to -1, then all rows of A add to zero. The equation $A\boldsymbol{x} = \mathbf{0}$ will now have the nonzero solution $\boldsymbol{x} = (1, 1, 1)$. A clear signal: ***This new A can't be inverted.***

2.5 B Three of these matrices are invertible, and three are singular. Find the inverse when it exists. Give reasons for noninvertibility (zero determinant, too few pivots, nonzero solution to $A\boldsymbol{x} = \mathbf{0}$) for the other three. The matrices are in the order A, B, C, D, S, E:

$$\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \quad \begin{bmatrix} 4 & 3 \\ 8 & 7 \end{bmatrix} \quad \begin{bmatrix} 6 & 6 \\ 6 & 0 \end{bmatrix} \quad \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Solution

$$B^{-1} = \frac{1}{4}\begin{bmatrix} 7 & -3 \\ -8 & 4 \end{bmatrix} \quad C^{-1} = \frac{1}{36}\begin{bmatrix} 0 & 6 \\ 6 & -6 \end{bmatrix} \quad S^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

A is not invertible because its determinant is $4 \cdot 6 - 3 \cdot 8 = 24 - 24 = 0$. D is not invertible because there is only one pivot; the second row becomes zero when the first row is subtracted. E is not invertible because a combination of the columns (the second column minus the first column) is zero—in other words $E\boldsymbol{x} = \mathbf{0}$ has the solution $\boldsymbol{x} = (-1, 1, 0)$.

Of course all three reasons for noninvertibility would apply to each of A, D, E.

2.5 C Apply the Gauss-Jordan method to invert this triangular "Pascal matrix" L. You see **Pascal's triangle**—adding each entry to the entry on its left gives the entry below. The entries of L are "binomial coefficients". The next row would be $1, 4, 6, 4, 1$.

$$\textbf{Triangular Pascal matrix} \quad L = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} \end{bmatrix} = \text{abs(pascal (4,1))}$$

Solution Gauss-Jordan starts with $[\,L \;\; I\,]$ and produces zeros by subtracting row 1:

$$[\,L \;\; I\,] = \left[\begin{array}{cccc|cccc} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & \mathbf{-1} & 1 & 0 & 0 \\ \mathbf{0} & 2 & 1 & 0 & \mathbf{-1} & 0 & 1 & 0 \\ \mathbf{0} & 3 & 3 & 1 & \mathbf{-1} & 0 & 0 & 1 \end{array}\right].$$

The next stage creates zeros below the second pivot, using multipliers 2 and 3. Then the last stage subtracts 3 times the new row 3 from the new row 4:

$$\to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & \mathbf{0} & 1 & 0 & \mathbf{1} & \mathbf{-2} & 1 & 0 \\ 0 & \mathbf{0} & 3 & 1 & \mathbf{2} & \mathbf{-3} & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \mathbf{-1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 1 & 0 & \mathbf{1} & \mathbf{-2} & \mathbf{1} & 0 \\ 0 & 0 & 0 & 1 & \mathbf{-1} & \mathbf{3} & \mathbf{-3} & \mathbf{1} \end{array}\right] = [\,I \;\; L^{-1}\,].$$

All the pivots were 1! So we didn't need to divide rows by pivots to get I. The inverse matrix L^{-1} looks like L itself, except odd-numbered diagonals have minus signs.

The same pattern continues to n by n Pascal matrices, L^{-1} has "alternating diagonals".

Problem Set 2.5

1 Find the inverses (directly or from the 2 by 2 formula) of A, B, C:

$$A = \begin{bmatrix} 0 & 3 \\ 4 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 0 \\ 4 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 3 & 4 \\ 5 & 7 \end{bmatrix}.$$

2 For these "permutation matrices" find P^{-1} by trial and error (with 1's and 0's):

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

3 Solve for the first column (x, y) and second column (t, z) of A^{-1}:

$$\begin{bmatrix} 10 & 20 \\ 20 & 50 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 10 & 20 \\ 20 & 50 \end{bmatrix} \begin{bmatrix} t \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

4 Show that $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 6 \end{smallmatrix}\right]$ is not invertible by trying to solve $AA^{-1} = I$ for column 1 of A^{-1}:

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \left(\begin{array}{l} \textit{For a different } A, \text{ could column 1 of } A^{-1} \\ \text{be possible to find but not column 2?} \end{array}\right)$$

5 Find an upper triangular U (not diagonal) with $U^2 = I$ which gives $U = U^{-1}$.

6 (a) If A is invertible and $AB = AC$, prove quickly that $B = C$.
(b) If $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$, find two different matrices such that $AB = AC$.

7 (Important) If A has row 1 + row 2 = row 3, show that A is not invertible:

(a) Explain why $A\boldsymbol{x} = (1, 0, 0)$ cannot have a solution.
(b) Which right sides (b_1, b_2, b_3) might allow a solution to $A\boldsymbol{x} = \boldsymbol{b}$?
(c) What happens to row 3 in elimination?

8 If A has column 1 + column 2 = column 3, show that A is not invertible:

(a) Find a nonzero solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{0}$. The matrix is 3 by 3.
(b) Elimination keeps column 1 + column 2 = column 3. Explain why there is no third pivot.

9 Suppose A is invertible and you exchange its first two rows to reach B. Is the new matrix B invertible and how would you find B^{-1} from A^{-1}?

10 Find the inverses (in any legal way) of

$$A = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 0 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 4 & 3 & 0 & 0 \\ 0 & 0 & 6 & 5 \\ 0 & 0 & 7 & 6 \end{bmatrix}.$$

11 (a) Find invertible matrices A and B such that $A+B$ is not invertible.

(b) Find singular matrices A and B such that $A+B$ is invertible.

12 If the product $C = AB$ is invertible (A and B are square), then A itself is invertible. Find a formula for A^{-1} that involves C^{-1} and B.

13 If the product $M = ABC$ of three square matrices is invertible, then B is invertible. (So are A and C.) Find a formula for B^{-1} that involves M^{-1} and A and C.

14 If you add row 1 of A to row 2 to get B, how do you find B^{-1} from A^{-1}?

$$\text{Notice the order.}\quad \text{The inverse of} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} & A & \end{bmatrix} \quad \text{is} \quad \underline{\qquad}.$$

15 Prove that a matrix with a column of zeros cannot have an inverse.

16 Multiply $\left[\begin{smallmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{smallmatrix}\right]$ times $\left[\begin{smallmatrix} \mathbf{d} & \mathbf{-b} \\ \mathbf{-c} & \mathbf{a} \end{smallmatrix}\right]$. What is the inverse of each matrix if $ad \neq bc$?

17 (a) What matrix E has the same effect as these three steps? Subtract row 1 from row 2, subtract row 1 from row 3, then subtract row 2 from row 3.

(b) What single matrix L has the same effect as these three reverse steps? Add row 2 to row 3, add row 1 to row 3, then add row 1 to row 2.

18 If B is the inverse of A^2, show that AB is the inverse of A.

19 Find the numbers a and b that give the inverse of $5 * \text{eye}(4) - \text{ones}(4,4)$:

$$\begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}.$$

What are a and b in the inverse of $6 * \text{eye}(5) - \text{ones}(5,5)$?

20 Show that $\text{A} = 4 * \text{eye}(4) - \text{ones}(4,4)$ is *not* invertible: Multiply $\text{A} * \text{ones}(4,1)$.

21 There are sixteen 2 by 2 matrices whose entries are 1's and 0's. How many of them are invertible?

Questions 22–28 are about the Gauss-Jordan method for calculating A^{-1}.

22 Change I into A^{-1} as you reduce A to I (by row operations):

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 & 0 \\ 2 & 7 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 3 & 9 & 0 & 1 \end{bmatrix}$$

23 Follow the 3 by 3 text example but with plus signs in A. Eliminate above and below the pivots to reduce $[\,A \quad I\,]$ to $[\,I \quad A^{-1}\,]$:

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

24 Use Gauss-Jordan elimination on $[\,U \quad I\,]$ to find the upper triangular U^{-1}:

$$\boldsymbol{U}\boldsymbol{U}^{-1} = \boldsymbol{I} \qquad \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

25 Find A^{-1} and B^{-1} (*if they exist*) by elimination on $[\,A \quad I\,]$ and $[\,B \quad I\,]$:

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

26 What three matrices E_{21} and E_{12} and D^{-1} reduce $A = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 6 \end{smallmatrix}\right]$ to the identity matrix? Multiply $D^{-1}E_{12}E_{21}$ to find A^{-1}.

27 Invert these matrices A by the Gauss-Jordan method starting with $[\,A \quad I\,]$:

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

28 Exchange rows and continue with Gauss-Jordan to find A^{-1}:

$$[\,A \quad I\,] = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix}.$$

29 True or false (with a counterexample if false and a reason if true):

(a) A 4 by 4 matrix with a row of zeros is not invertible.

(b) Every matrix with 1's down the main diagonal is invertible.

(c) If A is invertible then A^{-1} and A^2 are invertible.

30 For which three numbers c is this matrix not invertible, and why not?

$$A = \begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix}.$$

31 Prove that A is invertible if $a \neq 0$ and $a \neq b$ (find the pivots or A^{-1}):

$$A = \begin{bmatrix} a & b & b \\ a & a & b \\ a & a & a \end{bmatrix}.$$

32 This matrix has a remarkable inverse. Find A^{-1} by elimination on $[\,A \quad I\,]$. Extend to a 5 by 5 "alternating matrix" and guess its inverse; then multiply to confirm.

$$\text{Invert } A = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and solve } A\boldsymbol{x} = (1, 1, 1, 1).$$

33 Suppose the matrices P and Q have the same rows as I but in any order. They are "permutation matrices". Show that $P - Q$ is singular by solving $(P - Q)\boldsymbol{x} = \mathbf{0}$.

34 Find and check the inverses (assuming they exist) of these block matrices:

$$\begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \quad \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \quad \begin{bmatrix} 0 & I \\ I & D \end{bmatrix}.$$

35 Could a 4 by 4 matrix A be invertible if every row contains the numbers $0, 1, 2, 3$ in some order? What if every row of B contains $0, 1, 2, -3$ in some order?

36 In the Worked Example **2.5 C**, the triangular Pascal matrix L has an inverse with "alternating diagonals". Check that this L^{-1} is DLD, where the diagonal matrix D has alternating entries $1, -1, 1, -1$. Then $LDLD = I$, so what is the inverse of $LD =$ pascal (4,1)?

37 The Hilbert matrices have $H_{ij} = 1/(i + j - 1)$. Ask MATLAB for the exact 6 by 6 inverse invhilb(6). Then ask it to compute inv(hilb(6)). How can these be different, when the computer never makes mistakes?

38 (a) Use inv(P) to invert MATLAB's 4 by 4 symmetric matrix $P =$ pascal(4).

(b) Create Pascal's lower triangular $L =$ abs(pascal(4,1)) and test $P = LL^{\mathrm{T}}$.

39 If $A =$ ones(4) and $\boldsymbol{b} =$ rand(4,1), how does MATLAB tell you that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution? For the special $\boldsymbol{b} =$ ones(4,1), which solution to $A\boldsymbol{x} = \boldsymbol{b}$ is found by $A \backslash \boldsymbol{b}$?

Challenge Problems

40 (Recommended) A is a 4 by 4 matrix with 1's on the diagonal and $-a, -b, -c$ on the diagonal above. Find A^{-1} for this bidiagonal matrix.

41 Suppose E_1, E_2, E_3 are 4 by 4 identity matrices, except E_1 has a, b, c in column 1 and E_2 has d, e in column 2 and E_3 has f in column 3 (below the 1's). Multiply $L = E_1 E_2 E_3$ to show that all these nonzeros are copied into L.

$E_1 E_2 E_3$ is in the *opposite* order from elimination (because E_3 is acting first). But $E_1 E_2 E_3 = L$ is in the *correct* order to invert elimination and recover A.

42 Direct multiplications **1–4** give $MM^{-1} = I$, and I would recommend doing #**3**. M^{-1} shows the change in A^{-1} (useful to know) when a matrix is subtracted from A:

$$\begin{array}{llll}
\mathbf{1} & M = I - \boldsymbol{uv} & \text{and} & M^{-1} = I + \boldsymbol{uv}/(1 - \boldsymbol{vu}) \quad \text{(rank 1 change in } I)\\
\mathbf{2} & M = A - \boldsymbol{uv} & \text{and} & M^{-1} = A^{-1} + A^{-1}\boldsymbol{uv}A^{-1}/(1 - \boldsymbol{v}A^{-1}\boldsymbol{u})\\
\mathbf{3} & M = I - UV & \text{and} & M^{-1} = I_n + U(I_m - VU)^{-1}V\\
\mathbf{4} & M = A - UW^{-1}V & \text{and} & M^{-1} = A^{-1} + A^{-1}U(W - VA^{-1}U)^{-1}VA^{-1}
\end{array}$$

The Woodbury-Morrison formula **4** is the "matrix inversion lemma" in engineering. The ***Kalman filter*** for solving block tridiagonal systems uses formula **4** at each step. The four matrices M^{-1} are in diagonal blocks when inverting these block matrices ($\boldsymbol{v}$ is 1 by n, $\boldsymbol{u}$ is n by 1, V is m by n, U is n by m).

$$\begin{bmatrix} I & \boldsymbol{u} \\ \boldsymbol{v} & 1 \end{bmatrix} \qquad \begin{bmatrix} A & \boldsymbol{u} \\ \boldsymbol{v} & 1 \end{bmatrix} \qquad \begin{bmatrix} I_n & U \\ V & I_m \end{bmatrix} \qquad \begin{bmatrix} A & U \\ V & W \end{bmatrix}$$

43 Second difference matrices have beautiful inverses if they start with $T_{11} = 1$ (instead of $K_{11} = 2$). Here is the 3 by 3 tridiagonal matrix T and its inverse:

$$\boldsymbol{T_{11} = 1} \qquad T = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

One approach is Gauss-Jordan elimination on $[\,T \quad I\,]$. That seems too mechanical. I would rather write T as the product of first differences L times U. The inverses of L and U in Worked Example **2.5 A** are **sum matrices**, so here are T and T^{-1}:

$$LU = \underset{\textbf{difference}}{\begin{bmatrix} 1 & & \\ -1 & 1 & \\ 0 & -1 & 1 \end{bmatrix}} \underset{\textbf{difference}}{\begin{bmatrix} 1 & -1 & 0 \\ & 1 & -1 \\ & & 1 \end{bmatrix}} \qquad U^{-1}L^{-1} = \underset{\textbf{sum}}{\begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}} \underset{\textbf{sum}}{\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix}}$$

Question. (**4 by 4**) What are the pivots of T? What is its 4 by 4 inverse? The reverse order UL gives what matrix T^*? What is the inverse of T^*?

44 Here are two more difference matrices, both important. ***But are they invertible***?

$$\textbf{Cyclic } C = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \quad \textbf{Free ends } F = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

One test is elimination—the fourth pivot fails. Another test is the determinant, we don't want that. The best way is much faster, and independent of matrix size:

Produce $\boldsymbol{x} \neq \mathbf{0}$ so that $C\boldsymbol{x} = \mathbf{0}$. Do the same for $F\boldsymbol{x} = \mathbf{0}$. Not invertible.

Show how both equations $C\boldsymbol{x} = \boldsymbol{b}$ and $F\boldsymbol{x} = \boldsymbol{b}$ lead to $0 = b_1 + b_2 + \cdots + b_n$. There is no solution for other $\boldsymbol{b}$.

45 *Elimination for a* 2 *by* 2 *block matrix*: When you multiply the first block row by CA^{-1} and subtract from the second row, the "*Schur complement*" S appears:

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & S \end{bmatrix} \qquad \begin{matrix} A \text{ and } D \text{ are square} \\ S = D - CA^{-1}B. \end{matrix}$$

Multiply on the right to subtract $A^{-1}B$ times block column 1 from block column 2.

$$\begin{bmatrix} A & B \\ 0 & S \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = ? \quad \text{Find } S \text{ for} \quad \begin{bmatrix} A & B \\ C & I \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 \\ 4 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}.$$

The block pivots are A ***and*** S***. If they are invertible, so is*** $[\,A \quad B\,;\ \ C \quad D\,]$.

46 How does the identity $A(I + BA) = (I + AB)A$ connect the inverses of $I + BA$ and $I + AB$? Those are both invertible or both singular: not obvious.

1 Eigenvalues and Eigenvectors

The product $\mathbf{Ax}$ of a matrix $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and an n-vector $\mathbf{x}$ is itself an n-vector. Of particular interest in many settings (of which differential equations is one) is the following question:

> For a given matrix $\mathbf{A}$, what are the vectors $\mathbf{x}$ for which the product $\mathbf{Ax}$ is a scalar multiple of $\mathbf{x}$? That is, what vectors $\mathbf{x}$ satisfy the equation
>
> $$\mathbf{Ax} = \lambda\mathbf{x}$$
>
> for some scalar λ?

It should immediately be clear that, no matter what $\mathbf{A}$ and λ are, the vector $\mathbf{x} = \mathbf{0}$ (that is, the vector whose elements are all zero) satisfies this equation. With such a trivial answer, we might ask the question again in another way:

> For a given matrix $\mathbf{A}$, what are the *nonzero* vectors $\mathbf{x}$ that satisfy the equation
>
> $$\mathbf{Ax} = \lambda\mathbf{x}$$
>
> for some scalar λ?

To answer this question, we first perform some algebraic manipulations upon the equation $\mathbf{Ax} = \lambda\mathbf{x}$. We note first that, if $\mathbf{I} = \mathbf{I_n}$ (the $n \times n$ multiplicative identity in $\mathcal{M}_{n\times n}(\mathbb{R})$), then we can write

$$\begin{aligned}\mathbf{Ax} = \lambda\mathbf{x} \quad &\Leftrightarrow \quad \mathbf{Ax} - \lambda\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \quad \mathbf{Ax} - \lambda\mathbf{Ix} = \mathbf{0} \\ &\Leftrightarrow \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.\end{aligned}$$

Remember that we are looking for nonzero $\mathbf{x}$ that satisfy this last equation. But $\mathbf{A} - \lambda\mathbf{I}$ is an $n \times n$ matrix and, should its determinant be nonzero, this last equation will have exactly one solution, namely $\mathbf{x} = \mathbf{0}$. Thus our question above has the following answer:

> The equation $\mathbf{Ax} = \lambda\mathbf{x}$ has nonzero solutions for the vector x if and only if the matrix $\mathbf{A} - \lambda\mathbf{I}$ has zero determinant.

As we will see in the examples below, for a given matrix $\mathbf{A}$ there are only a few special values of the scalar λ for which $\mathbf{A} - \lambda\mathbf{I}$ will have zero determinant, and these special values are called the *eigenvalues* of the matrix $\mathbf{A}$. Based upon the answer to our question, it seems we must first be able to find the eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_n$ of $\mathbf{A}$ and then see about solving the individual equations $\mathbf{Ax} = \lambda_i\mathbf{x}$ for each $i = 1, \ldots, n$.

Example: Find the eigenvalues of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix}$.

The eigenvalues are those λ for which $\det(\mathbf{A}-\lambda\mathbf{I})= 0$. Now

$$\begin{aligned}
\det(\mathbf{A}-\lambda\mathbf{I}) &= \det\left(\begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\
&= \begin{vmatrix} 2-\lambda & 2 \\ 5 & -1-\lambda \end{vmatrix} \\
&= (2-\lambda)(-1-\lambda) - 10 \\
&= \lambda^2 - \lambda - 12.
\end{aligned}$$

The eigenvalues of $\mathbf{A}$ are the solutions of the quadratic equation $\lambda^2 - \lambda - 12 = 0$, namely $\lambda_1 = -3$ and $\lambda_2 = 4$.

As we have discussed, if $\det(\mathbf{A}-\lambda\mathbf{I}) = 0$ then the equation $(\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = \mathbf{b}$ has either no solutions or infinitely many. When we take $\mathbf{b} = \mathbf{0}$ however, it is clear by the existence of the solution $\mathbf{x} = \mathbf{0}$ that there are infinitely many solutions (i.e., we may rule out the "no solution" case). If we continue using the matrix $\mathbf{A}$ from the example above, we can expect nonzero solutions $\mathbf{x}$ (infinitely many of them, in fact) of the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ precisely when $\lambda = -3$ or $\lambda = 4$. Let us procede to characterize such solutions.

First, we work with $\lambda = -3$. The equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ becomes $\mathbf{A}\mathbf{x} = -3\mathbf{x}$. Writing

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and using the matrix $\mathbf{A}$ from above, we have

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix},$$

while

$$-3\mathbf{x} = \begin{bmatrix} -3x_1 \\ -3x_2 \end{bmatrix}.$$

Setting these equal, we get

$$\begin{aligned}
\begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix} = \begin{bmatrix} -3x_1 \\ -3x_2 \end{bmatrix} \quad &\Rightarrow \quad 2x_1 + 2x_2 = -3x_1 \qquad \text{and} \qquad 5x_1 - x_2 = -3x_2 \\
&\Rightarrow \quad 5x_1 = -2x_2 \\
&\Rightarrow \quad x_1 = -\frac{2}{5}x_2.
\end{aligned}$$

This means that, while there are infinitely many nonzero solutions (solution vectors) of the equation $\mathbf{A}\mathbf{x} = -3\mathbf{x}$, they all satisfy the condition that the first entry x_1 is $-2/5$ times the second entry x_2. Thus all solutions of this equation can be characterized by

$$\begin{bmatrix} 2t \\ -5t \end{bmatrix} = t\begin{bmatrix} 2 \\ -5 \end{bmatrix},$$

where t is any real number. The nonzero vectors $\mathbf{x}$ that satisfy $\mathbf{Ax} = -3\mathbf{x}$ are called *eigenvectors* associated with the eigenvalue $\lambda = -3$. One such eigenvector is

$$\mathbf{u_1} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$$

and all other eigenvectors corresponding to the eigenvalue (-3) are simply scalar multiples of $\mathbf{u_1}$ — that is, $\mathbf{u_1}$ spans this set of eigenvectors.

Similarly, we can find eigenvectors associated with the eigenvalue $\lambda = 4$ by solving $\mathbf{Ax} = 4\mathbf{x}$:

$$\begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 \\ 4x_2 \end{bmatrix} \quad \begin{aligned} &\Rightarrow \; 2x_1 + 2x_2 = 4x_1 \qquad \text{and} \qquad 5x_1 - x_2 = 4x_2 \\ &\Rightarrow \; x_1 = x_2. \end{aligned}$$

Hence the set of eigenvectors associated with $\lambda = 4$ is spanned by

$$\mathbf{u_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Example: Find the eigenvalues and associated eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix}.$$

First we compute $\det(\mathbf{A} - \lambda\mathbf{I})$ via a cofactor expansion along the second column:

$$\begin{aligned} \begin{vmatrix} 7-\lambda & 0 & -3 \\ -9 & -2-\lambda & 3 \\ 18 & 0 & -8-\lambda \end{vmatrix} &= (-2-\lambda)(-1)^4 \begin{vmatrix} 7-\lambda & -3 \\ 18 & -8-\lambda \end{vmatrix} \\ &= -(2+\lambda)[(7-\lambda)(-8-\lambda)+54] \\ &= -(\lambda+2)(\lambda^2+\lambda-2) \\ &= -(\lambda+2)^2(\lambda-1). \end{aligned}$$

Thus $\mathbf{A}$ has two distinct eigenvalues, $\lambda_1 = -2$ and $\lambda_3 = 1$. (Note that we might say $\lambda_2 = -2$, since, as a root, -2 has multiplicity two. This is why we labelled the eigenvalue 1 as λ_3.)

Now, to find the associated eigenvectors, we solve the equation $(\mathbf{A} - \lambda_j\mathbf{I})\mathbf{x} = \mathbf{0}$ for $j = 1, 2, 3$. Using the eigenvalue $\lambda_3 = 1$, we have

$$\begin{aligned} (\mathbf{A} - \mathbf{I})\mathbf{x} &= \begin{bmatrix} 6x_1 - 3x_3 \\ -9x_1 - 3x_2 + 3x_3 \\ 18x_1 - 9x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \; x_3 = 2x_1 \qquad \text{and} \qquad x_2 = x_3 - 3x_1 \\ &\Rightarrow \; x_3 = 2x_1 \qquad \text{and} \qquad x_2 = -x_1. \end{aligned}$$

So the eigenvectors associated with $\lambda_3 = 1$ are all scalar multiples of

$$\mathbf{u_3} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Now, to find eigenvectors associated with $\lambda_1 = -2$ we solve $(\mathbf{A} + 2\mathbf{I})\mathbf{x} = \mathbf{0}$. We have

$$\begin{aligned} (\mathbf{A} + 2\mathbf{I})\mathbf{x} &= \begin{bmatrix} 9x_1 - 3x_3 \\ -9x_1 + 3x_3 \\ 18x_1 - 6x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \; x_3 = 3x_1. \end{aligned}$$

Something different happened here in that we acquired no information about x_2. In fact, we have found that x_2 can be chosen arbitrarily, and independently of x_1 and x_3 (whereas x_3 cannot be chosen independently of x_1). This allows us to choose two linearly independent eigenvectors associated with the eigenvalue $\lambda = -2$, such as $\mathbf{u_1} = (1, 0, 3)$ and $\mathbf{u_2} = (1, 1, 3)$. It is a fact that all other eigenvectors associated with $\lambda_2 = -2$ are in the span of these two; that is, all others can be written as linear combinations $c_1\mathbf{u_1} + c_2\mathbf{u_2}$ using an appropriate choices of the constants c_1 and c_2.

Example: Find the eigenvalues and associated eigenvectors of the matrix

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

We compute

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= \begin{vmatrix} -1-\lambda & 2 \\ 0 & -1-\lambda \end{vmatrix} \\ &= (\lambda + 1)^2. \end{aligned}$$

Setting this equal to zero we get that $\lambda = -1$ is a (repeated) eigenvalue. To find any associated eigenvectors we must solve for $\mathbf{x} = (x_1, x_2)$ so that $(\mathbf{A} + \mathbf{I})\mathbf{x} = \mathbf{0}$; that is,

$$\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad x_2 = 0.$$

Thus, the eigenvectors corresponding to the eigenvalue $\lambda = -1$ are the vectors whose second component is zero, which means that we are talking about all scalar multiples of $\mathbf{u} = (1, 0)$.

Notice that our work above shows that there are no eigenvectors associated with $\lambda = -1$ which are linearly independent of $\mathbf{u}$. This may go against your intuition based upon the results of the example before this one, where an eigenvalue of multiplicity two had two linearly independent associated eigenvectors. Nevertheless, it is a (somewhat disparaging) fact that eigenvalues can have fewer linearly independent eigenvectors than their multiplicity suggests.

Example: Find the eigenvalues and associated eigenvectors of the matrix

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

We compute

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= \begin{vmatrix} 2-\lambda & -1 \\ 1 & 2-\lambda \end{vmatrix} \\ &= (\lambda-2)^2 + 1 \\ &= \lambda^2 - 4\lambda + 5. \end{aligned}$$

The roots of this polynomial are $\lambda_1 = 2+i$ and $\lambda_2 = 2-i$; that is, the eigenvalues are not real numbers. This is a common occurrence, and we can press on to find the eigenvectors just as we have in the past with real eigenvalues. To find eigenvectors associated with $\lambda_1 = 2+i$, we look for $\mathbf{x}$ satisfying

$$\begin{aligned} (\mathbf{A} - (2+i)\mathbf{I})\mathbf{x} = \mathbf{0} \quad &\Rightarrow \quad \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \quad \begin{bmatrix} -ix_1 - x_2 \\ x_1 - ix_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \quad x_1 = ix_2. \end{aligned}$$

Thus all eigenvectors associated with $\lambda_1 = 2+i$ are scalar multiples of $\mathbf{u_1} = (i, 1)$. Proceeding with $\lambda_2 = 2-i$, we have

$$\begin{aligned} (\mathbf{A} - (2-i)\mathbf{I})\mathbf{x} = \mathbf{0} \quad &\Rightarrow \quad \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \quad \begin{bmatrix} ix_1 - x_2 \\ x_1 + ix_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\Rightarrow \quad x_1 = -ix_2, \end{aligned}$$

which shows all eigenvectors associated with $\lambda_2 = 2-i$ to be scalar multiples of $\mathbf{u_2} = (-i, 1)$.

Notice that $\mathbf{u_2}$, the eigenvector associated with the eigenvalue $\lambda_2 = 2-i$ in the last example, is the complex conjugate of $\mathbf{u_1}$, the eigenvector associated with the eigenvalue $\lambda_1 = 2+i$. It is indeed a fact that, if $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ has a nonreal eigenvalue $\lambda_1 = \lambda + i\mu$ with corresponding eigenvector ξ_1, then it also has eigenvalue $\lambda_2 = \lambda - i\mu$ with corresponding eigenvector $\xi_2 = \bar{\xi}_1$.

6. Vector Spaces

In this chapter we introduce vector spaces in full generality. The reader will notice some similarity with the discussion of the space $\mathbb{R}^n$ in Chapter 5. In fact much of the present material has been developed in that context, and there is some repetition. However, Chapter 6 deals with the notion of an *abstract* vector space, a concept that will be new to most readers. It turns out that there are many systems in which a natural addition and scalar multiplication are defined and satisfy the usual rules familiar from $\mathbb{R}^n$. The study of abstract vector spaces is a way to deal with all these examples simultaneously. The new aspect is that we are dealing with an abstract system in which *all we know* about the vectors is that they are objects that can be added and multiplied by a scalar and satisfy rules familiar from $\mathbb{R}^n$.

The novel thing is the *abstraction*. Getting used to this new conceptual level is facilitated by the work done in Chapter 5: First, the vector manipulations are familiar, giving the reader more time to become accustomed to the abstract setting; and, second, the mental images developed in the concrete setting of $\mathbb{R}^n$ serve as an aid to doing many of the exercises in Chapter 6.

The concept of a vector space was first introduced in 1844 by the German mathematician Hermann Grassmann (1809-1877), but his work did not receive the attention it deserved. It was not until 1888 that the Italian mathematician Guiseppe Peano (1858-1932) clarified Grassmann's work in his book *Calcolo Geometrico* and gave the vector space axioms in their present form. Vector spaces became established with the work of the Polish mathematician Stephan Banach (1892-1945), and the idea was finally accepted in 1918 when Hermann Weyl (1885-1955) used it in his widely read book *Raum-Zeit-Materie* ("Space-Time-Matter"), an introduction to the general theory of relativity.

6.1 Examples and Basic Properties

Many mathematical entities have the property that they can be added and multiplied by a number. Numbers themselves have this property, as do $m \times n$ matrices: The sum of two such matrices is again $m \times n$ as is any scalar multiple of such a matrix. Polynomials are another familiar example, as are the geometric vectors in Chapter 4. It turns out that there are many other types of mathematical objects that can be added and multiplied by a scalar, and the general study of such systems is introduced in this chapter. Remarkably, much of what we could say in Chapter 5 about the dimension of subspaces in $\mathbb{R}^n$ can be formulated in this generality.

Definition 6.1 Vector Spaces

*A **vector space** consists of a nonempty set V of objects (called **vectors**) that can be added, that can be multiplied by a real number (called a **scalar** in this context), and for which certain axioms hold.*[1] *If $\mathbf{v}$ and $\mathbf{w}$ are two vectors in V, their sum is expressed as $\mathbf{v}+\mathbf{w}$, and the scalar product of $\mathbf{v}$ by a real number a is denoted as $a\mathbf{v}$. These operations are called **vector addition** and **scalar multiplication**, respectively, and the following axioms are assumed to hold.*

Axioms for vector addition

A1. *If $\mathbf{u}$ and $\mathbf{v}$ are in V, then $\mathbf{u}+\mathbf{v}$ is in V.*

A2. *$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ for all $\mathbf{u}$ and $\mathbf{v}$ in V.*

A3. *$\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$ for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V.*

A4. *An element $\mathbf{0}$ in V exists such that $\mathbf{v}+\mathbf{0}=\mathbf{v}=\mathbf{0}+\mathbf{v}$ for every $\mathbf{v}$ in V.*

A5. *For each $\mathbf{v}$ in V, an element $-\mathbf{v}$ in V exists such that $-\mathbf{v}+\mathbf{v}=\mathbf{0}$ and $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$.*

Axioms for scalar multiplication

S1. *If $\mathbf{v}$ is in V, then $a\mathbf{v}$ is in V for all a in $\mathbb{R}$.*

S2. *$a(\mathbf{v}+\mathbf{w})=a\mathbf{v}+a\mathbf{w}$ for all $\mathbf{v}$ and $\mathbf{w}$ in V and all a in $\mathbb{R}$.*

S3. *$(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$ for all $\mathbf{v}$ in V and all a and b in $\mathbb{R}$.*

S4. *$a(b\mathbf{v})=(ab)\mathbf{v}$ for all $\mathbf{v}$ in V and all a and b in $\mathbb{R}$.*

S5. *$1\mathbf{v}=\mathbf{v}$ for all $\mathbf{v}$ in V.*

The content of axioms A1 and S1 is described by saying that V is **closed** under vector addition and scalar multiplication. The element $\mathbf{0}$ in axiom A4 is called the **zero vector**, and the vector $-\mathbf{v}$ in axiom A5 is called the **negative** of $\mathbf{v}$.

The rules of matrix arithmetic, when applied to $\mathbb{R}^n$, give

Example 6.1.1

$\mathbb{R}^n$ is a vector space using matrix addition and scalar multiplication.[2]

It is important to realize that, in a general vector space, the vectors need not be n-tuples as in $\mathbb{R}^n$. They can be any kind of objects at all as long as the addition and scalar multiplication are defined and the axioms are satisfied. The following examples illustrate the diversity of the concept.

The space $\mathbb{R}^n$ consists of special types of matrices. More generally, let $\mathbf{M}_{mn}$ denote the set of all $m\times n$ matrices with real entries. Then Theorem 2.1.1 gives:

[1]The scalars will usually be real numbers, but they could be complex numbers, or elements of an algebraic system called a field. Another example is the field $\mathbb{Q}$ of rational numbers. We will look briefly at finite fields in Section 8.8.

[2]We will usually write the vectors in $\mathbb{R}^n$ as n-tuples. However, if it is convenient, we will sometimes denote them as rows or columns.

Example 6.1.2

The set $\mathbf{M}_{mn}$ of all $m \times n$ matrices is a vector space using matrix addition and scalar multiplication. The zero element in this vector space is the zero matrix of size $m \times n$, and the vector space negative of a matrix (required by axiom A5) is the usual matrix negative discussed in Section 2.1. Note that $\mathbf{M}_{mn}$ is just $\mathbb{R}^{mn}$ in different notation.

In Chapter 5 we identified many important subspaces of $\mathbb{R}^n$ such as $\operatorname{im} A$ and $\operatorname{null} A$ for a matrix A. These are all vector spaces.

Example 6.1.3

Show that every subspace of $\mathbb{R}^n$ is a vector space in its own right using the addition and scalar multiplication of $\mathbb{R}^n$.

Solution. Axioms A1 and S1 are two of the defining conditions for a subspace U of $\mathbb{R}^n$ (see Section 5.1). The other eight axioms for a vector space are inherited from $\mathbb{R}^n$. For example, if $\mathbf{x}$ and $\mathbf{y}$ are in U and a is a scalar, then $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x}+a\mathbf{y}$ because $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$. This shows that axiom S2 holds for U; similarly, the other axioms also hold for U.

Example 6.1.4

Let V denote the set of all ordered pairs (x, y) and define addition in V as in $\mathbb{R}^2$. However, define a new scalar multiplication in V by

$$a(x, y) = (ay, ax)$$

Determine if V is a vector space with these operations.

Solution. Axioms A1 to A5 are valid for V because they hold for matrices. Also $a(x, y) = (ay, ax)$ is again in V, so axiom S1 holds. To verify axiom S2, let $\mathbf{v} = (x, y)$ and $\mathbf{w} = (x_1, y_1)$ be typical elements in V and compute

$$\begin{aligned} a(\mathbf{v}+\mathbf{w}) &= a(x+x_1, y+y_1) = (a(y+y_1), a(x+x_1)) \\ a\mathbf{v}+a\mathbf{w} &= (ay, ax)+(ay_1, ax_1) = (ay+ay_1, ax+ax_1) \end{aligned}$$

Because these are equal, axiom S2 holds. Similarly, the reader can verify that axiom S3 holds. However, axiom S4 fails because

$$a(b(x, y)) = a(by, bx) = (abx, aby)$$

need not equal $ab(x, y) = (aby, abx)$. Hence, V is *not* a vector space. (In fact, axiom S5 also fails.)

Sets of polynomials provide another important source of examples of vector spaces, so we review some basic facts. A **polynomial** in an indeterminate x is an expression

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0,\ a_1,\ a_2,\ \dots,\ a_n$ are real numbers called the **coefficients** of the polynomial. If all the coefficients are zero, the polynomial is called the **zero polynomial** and is denoted simply as 0. If $p(x) \neq 0$, the highest power of x with a nonzero coefficient is called the **degree** of $p(x)$ denoted as $\deg p(x)$. The coefficient itself is called the **leading coefficient** of $p(x)$. Hence $\deg(3+5x) = 1$, $\deg(1+x+x^2) = 2$, and $\deg(4) = 0$. (The degree of the zero polynomial is not defined.)

Let **P** denote the set of all polynomials and suppose that

$$\begin{aligned} p(x) &= a_0 + a_1x + a_2x^2 + \cdots \\ q(x) &= b_0 + b_1x + b_2x^2 + \cdots \end{aligned}$$

are two polynomials in **P** (possibly of different degrees). Then $p(x)$ and $q(x)$ are called **equal** [written $p(x) = q(x)$] if and only if all the corresponding coefficients are equal—that is, $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$, and so on. In particular, $a_0 + a_1x + a_2x^2 + \cdots = 0$ means that $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, $\dots$, and this is the reason for calling x an **indeterminate**. The set **P** has an addition and scalar multiplication defined on it as follows: if $p(x)$ and $q(x)$ are as before and a is a real number,

$$\begin{aligned} p(x) + q(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots \\ ap(x) &= aa_0 + (aa_1)x + (aa_2)x^2 + \cdots \end{aligned}$$

Evidently, these are again polynomials, so **P** is closed under these operations, called **pointwise** addition and scalar multiplication. The other vector space axioms are easily verified, and we have

Example 6.1.5

The set **P** of all polynomials is a vector space with the foregoing addition and scalar multiplication. The zero vector is the zero polynomial, and the negative of a polynomial $p(x) = a_0 + a_1x + a_2x^2 + \dots$ is the polynomial $-p(x) = -a_0 - a_1x - a_2x^2 - \dots$ obtained by negating all the coefficients.

There is another vector space of polynomials that will be referred to later.

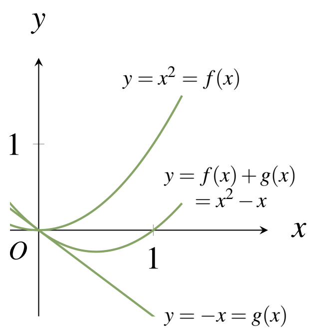
Example 6.1.6

Given $n \geq 1$, let $\mathbf{P}_n$ denote the set of all polynomials of degree at most n, together with the zero polynomial. That is

$$\mathbf{P}_n = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid a_0,\ a_1,\ a_2,\ \dots,\ a_n \text{ in } \mathbb{R}\}.$$

Then $\mathbf{P}_n$ is a vector space. Indeed, sums and scalar multiples of polynomials in $\mathbf{P}_n$ are again in $\mathbf{P}_n$, and the other vector space axioms are inherited from **P**. In particular, the zero vector and the negative of a polynomial in $\mathbf{P}_n$ are the same as those in **P**.

If a and b are real numbers and $a < b$, the **interval** $[a,\ b]$ is defined to be the set of all real numbers x such that $a \leq x \leq b$. A (real-valued) **function** f on $[a,\ b]$ is a rule that associates to every number x in $[a,\ b]$ a real number denoted $f(x)$. The rule is frequently specified by giving a formula for $f(x)$ in terms of x. For example, $f(x) = 2^x$, $f(x) = \sin x$, and $f(x) = x^2 + 1$ are familiar functions. In fact, every polynomial $p(x)$ can be regarded as the formula for a function p.



The set of all functions on $[a, b]$ is denoted $\mathbf{F}[a, b]$. Two functions f and g in $\mathbf{F}[a, b]$ are **equal** if $f(x) = g(x)$ for every x in $[a, b]$, and we describe this by saying that f and g have the **same action**. Note that two polynomials are equal in $\mathbf{P}$ (defined prior to Example 6.1.5) if and only if they are equal as functions.

If f and g are two functions in $\mathbf{F}[a, b]$, and if r is a real number, define the sum $f+g$ and the scalar product rf by

$$\begin{aligned} (f+g)(x) &= f(x)+g(x) \quad \text{for each } x \text{ in } [a, b] \\ (rf)(x) &= rf(x) \qquad\qquad \text{for each } x \text{ in } [a, b] \end{aligned}$$

In other words, the action of $f+g$ upon x is to associate x with the number $f(x)+g(x)$, and rf associates x with $rf(x)$. The sum of $f(x) = x^2$ and $g(x) = -x$ is shown in the diagram. These operations on $\mathbf{F}[a, b]$ are called **pointwise addition and scalar multiplication** of functions and they are the usual operations familiar from elementary algebra and calculus.

Example 6.1.7

The set $\mathbf{F}[a, b]$ of all functions on the interval $[a, b]$ is a vector space using pointwise addition and scalar multiplication. The zero function (in axiom A4), denoted 0, is the constant function defined by

$$0(x) = 0 \quad \text{for each } x \text{ in } [a, b]$$

The negative of a function f is denoted $-f$ and has action defined by

$$(-f)(x) = -f(x) \quad \text{for each } x \text{ in } [a, b]$$

Axioms A1 and S1 are clearly satisfied because, if f and g are functions on $[a, b]$, then $f+g$ and rf are again such functions. The verification of the remaining axioms is left as Exercise 6.1.14.

Other examples of vector spaces will appear later, but these are sufficiently varied to indicate the scope of the concept and to illustrate the properties of vector spaces to be discussed. With such a variety of examples, it may come as a surprise that a well-developed *theory* of vector spaces exists. That is, many properties can be shown to hold for *all* vector spaces and hence hold in every example. Such properties are called *theorems* and can be deduced from the axioms. Here is an important example.

Theorem 6.1.1: Cancellation

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in a vector space V. If $\mathbf{v}+\mathbf{u} = \mathbf{v}+\mathbf{w}$, then $\mathbf{u} = \mathbf{w}$.

Proof. We are given $\mathbf{v}+\mathbf{u} = \mathbf{v}+\mathbf{w}$. If these were numbers instead of vectors, we would simply subtract $\mathbf{v}$ from both sides of the equation to obtain $\mathbf{u} = \mathbf{w}$. This can be accomplished with vectors by adding $-\mathbf{v}$ to both sides of the equation. The steps (using only the axioms) are as follows:

$$\begin{aligned} \mathbf{v}+\mathbf{u} &= \mathbf{v}+\mathbf{w} \\ -\mathbf{v}+(\mathbf{v}+\mathbf{u}) &= -\mathbf{v}+(\mathbf{v}+\mathbf{w}) && \text{(axiom A5)} \\ (-\mathbf{v}+\mathbf{v})+\mathbf{u} &= (-\mathbf{v}+\mathbf{v})+\mathbf{w} && \text{(axiom A3)} \end{aligned}$$

$$\mathbf{0}+\mathbf{u}=\mathbf{0}+\mathbf{w} \quad \text{(axiom A5)}$$
$$\mathbf{u}=\mathbf{w} \quad \text{(axiom A4)}$$

This is the desired conclusion.[3] □

As with many good mathematical theorems, the technique of the proof of Theorem 6.1.1 is at least as important as the theorem itself. The idea was to mimic the well-known process of numerical subtraction in a vector space V as follows: To subtract a vector $\mathbf{v}$ from both sides of a vector equation, we added $-\mathbf{v}$ to both sides. With this in mind, we define **difference** $\mathbf{u}-\mathbf{v}$ of two vectors in V as

$$\mathbf{u}-\mathbf{v}=\mathbf{u}+(-\mathbf{v})$$

We shall say that this vector is the result of having **subtracted** $\mathbf{v}$ from $\mathbf{u}$ and, as in arithmetic, this operation has the property given in Theorem 6.1.2.

Theorem 6.1.2

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a vector space V, the equation

$$\mathbf{x}+\mathbf{v}=\mathbf{u}$$

has one and only one solution $\mathbf{x}$ in V given by

$$\mathbf{x}=\mathbf{u}-\mathbf{v}$$

Proof. The difference $\mathbf{x}=\mathbf{u}-\mathbf{v}$ is indeed a solution to the equation because (using several axioms)

$$\mathbf{x}+\mathbf{v}=(\mathbf{u}-\mathbf{v})+\mathbf{v}=[\mathbf{u}+(-\mathbf{v})]+\mathbf{v}=\mathbf{u}+(-\mathbf{v}+\mathbf{v})=\mathbf{u}+\mathbf{0}=\mathbf{u}$$

To see that this is the only solution, suppose $\mathbf{x}_1$ is another solution so that $\mathbf{x}_1+\mathbf{v}=\mathbf{u}$. Then $\mathbf{x}+\mathbf{v}=\mathbf{x}_1+\mathbf{v}$ (they both equal $\mathbf{u}$), so $\mathbf{x}=\mathbf{x}_1$ by cancellation. □

Similarly, cancellation shows that there is only one zero vector in any vector space and only one negative of each vector (Exercises 6.1.10 and 6.1.11). Hence we speak of *the* zero vector and *the* negative of a vector.

The next theorem derives some basic properties of scalar multiplication that hold in every vector space, and will be used extensively.

Theorem 6.1.3

Let $\mathbf{v}$ denote a vector in a vector space V and let a denote a real number.

1. $0\mathbf{v}=\mathbf{0}$.
2. $a\mathbf{0}=\mathbf{0}$.
3. *If $a\mathbf{v}=\mathbf{0}$, then either $a=0$ or $\mathbf{v}=\mathbf{0}$.*

[3] Observe that none of the scalar multiplication axioms are needed here.

4. $(-1)\mathbf{v} = -\mathbf{v}$.

5. $(-a)\mathbf{v} = -(a\mathbf{v}) = a(-\mathbf{v})$.

Proof.

1. Observe that $0\mathbf{v} + 0\mathbf{v} = (0+0)\mathbf{v} = 0\mathbf{v} = 0\mathbf{v} + \mathbf{0}$ where the first equality is by axiom S3. It follows that $0\mathbf{v} = \mathbf{0}$ by cancellation.

2. The proof is similar to that of (1), and is left as Exercise 6.1.12(a).

3. Assume that $a\mathbf{v} = \mathbf{0}$. If $a = 0$, there is nothing to prove; if $a \neq 0$, we must show that $\mathbf{v} = \mathbf{0}$. But $a \neq 0$ means we can scalar-multiply the equation $a\mathbf{v} = \mathbf{0}$ by the scalar $\frac{1}{a}$. The result (using (2) and Axioms S5 and S4) is
$$\mathbf{v} = 1\mathbf{v} = \left(\tfrac{1}{a}a\right)\mathbf{v} = \tfrac{1}{a}(a\mathbf{v}) = \tfrac{1}{a}\mathbf{0} = \mathbf{0}$$

4. We have $-\mathbf{v} + \mathbf{v} = \mathbf{0}$ by axiom A5. On the other hand,
$$(-1)\mathbf{v} + \mathbf{v} = (-1)\mathbf{v} + 1\mathbf{v} = (-1+1)\mathbf{v} = 0\mathbf{v} = \mathbf{0}$$
using (1) and axioms S5 and S3. Hence $(-1)\mathbf{v} + \mathbf{v} = -\mathbf{v} + \mathbf{v}$ (because both are equal to $\mathbf{0}$), so $(-1)\mathbf{v} = -\mathbf{v}$ by cancellation.

5. The proof is left as Exercise 6.1.12.[4] □

The properties in Theorem 6.1.3 are familiar for matrices; the point here is that they hold in *every* vector space. It is hard to exaggerate the importance of this observation.

Axiom A3 ensures that the sum $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ is the same however it is formed, and we write it simply as $\mathbf{u} + \mathbf{v} + \mathbf{w}$. Similarly, there are different ways to form any sum $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n$, and Axiom A3 guarantees that they are all equal. Moreover, Axiom A2 shows that the order in which the vectors are written does not matter (for example: $\mathbf{u} + \mathbf{v} + \mathbf{w} + \mathbf{z} = \mathbf{z} + \mathbf{u} + \mathbf{w} + \mathbf{v}$).

Similarly, Axioms S2 and S3 extend. For example
$$a(\mathbf{u} + \mathbf{v} + \mathbf{w}) = a\left[\mathbf{u} + (\mathbf{v} + \mathbf{w})\right] = a\mathbf{u} + a(\mathbf{v} + \mathbf{w}) = a\mathbf{u} + a\mathbf{v} + a\mathbf{w}$$
for all a, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. Similarly $(a+b+c)\mathbf{v} = a\mathbf{v} + b\mathbf{v} + c\mathbf{v}$ hold for all values of a, b, c, and $\mathbf{v}$ (verify). More generally,
$$\begin{aligned} a(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n) &= a\mathbf{v}_1 + a\mathbf{v}_2 + \cdots + a\mathbf{v}_n \\ (a_1 + a_2 + \cdots + a_n)\mathbf{v} &= a_1\mathbf{v} + a_2\mathbf{v} + \cdots + a_n\mathbf{v} \end{aligned}$$
hold for all $n \geq 1$, all numbers a, $a_1, \dots, a_n$, and all vectors, $\mathbf{v}$, $\mathbf{v}_1, \dots, \mathbf{v}_n$. The verifications are by induction and are left to the reader (Exercise 6.1.13). These facts—together with the axioms, Theorem 6.1.3, and the definition of subtraction—enable us to simplify expressions involving sums of scalar multiples of vectors by collecting like terms, expanding, and taking out common factors. This has been discussed for the vector space of matrices in Section 2.1 (and for geometric vectors in Section 4.1); the manipulations in an arbitrary vector space are carried out in the same way. Here is an illustration.

Example 6.1.8

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a vector space V, simplify the expression

$$2(\mathbf{u}+3\mathbf{w})-3(2\mathbf{w}-\mathbf{v})-3[2(2\mathbf{u}+\mathbf{v}-4\mathbf{w})-4(\mathbf{u}-2\mathbf{w})]$$

Solution. The reduction proceeds as though $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ were matrices or variables.

$$\begin{aligned}&2(\mathbf{u}+3\mathbf{w})-3(2\mathbf{w}-\mathbf{v})-3[2(2\mathbf{u}+\mathbf{v}-4\mathbf{w})-4(\mathbf{u}-2\mathbf{w})]\\ &=2\mathbf{u}+6\mathbf{w}-6\mathbf{w}+3\mathbf{v}-3[4\mathbf{u}+2\mathbf{v}-8\mathbf{w}-4\mathbf{u}+8\mathbf{w}]\\ &=2\mathbf{u}+3\mathbf{v}-3[2\mathbf{v}]\\ &=2\mathbf{u}+3\mathbf{v}-6\mathbf{v}\\ &=2\mathbf{u}-3\mathbf{v}\end{aligned}$$

Condition (2) in Theorem 6.1.3 points to another example of a vector space.

Example 6.1.9

A set $\{\mathbf{0}\}$ with one element becomes a vector space if we define

$$\mathbf{0}+\mathbf{0}=\mathbf{0} \quad \text{and} \quad a\mathbf{0}=\mathbf{0} \text{ for all scalars } a.$$

The resulting space is called the **zero vector space** and is denoted $\{\mathbf{0}\}$.

The vector space axioms are easily verified for $\{\mathbf{0}\}$. In any vector space V, Theorem 6.1.3 shows that the zero subspace (consisting of the zero vector of V alone) is a copy of the zero vector space.

Exercises for 6.1

Exercise 6.1.1 Let V denote the set of ordered triples (x, y, z) and define addition in V as in $\mathbb{R}^3$. For each of the following definitions of scalar multiplication, decide whether V is a vector space.

a. $a(x, y, z)=(ax, y, az)$

b. $a(x, y, z)=(ax, 0, az)$

c. $a(x, y, z)=(0, 0, 0)$

d. $a(x, y, z)=(2ax, 2ay, 2az)$

Exercise 6.1.2 Are the following sets vector spaces with the indicated operations? If not, why not?

a. The set V of nonnegative real numbers; ordinary addition and scalar multiplication.

b. The set V of all polynomials of degree ≥ 3, together with 0; operations of $\mathbf{P}$.

c. The set of all polynomials of degree ≤ 3; operations of $\mathbf{P}$.

d. The set $\{1, x, x^2, \dots\}$; operations of $\mathbf{P}$.

e. The set V of all 2×2 matrices of the form $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$; operations of $\mathbf{M}_{22}$.

f. The set V of 2×2 matrices with equal column sums; operations of $\mathbf{M}_{22}$.

g. The set V of 2×2 matrices with zero determinant; usual matrix operations.

h. The set V of real numbers; usual operations.

i. The set V of complex numbers; usual addition and multiplication by a real number.

j. The set V of all ordered pairs $(x,\ y)$ with the addition of $\mathbb{R}^2$, but using scalar multiplication $a(x,\ y) = (ax,\ -ay)$.

k. The set V of all ordered pairs $(x,\ y)$ with the addition of $\mathbb{R}^2$, but using scalar multiplication $a(x,\ y) = (x,\ y)$ for all a in $\mathbb{R}$.

l. The set V of all functions $f : \mathbb{R} \to \mathbb{R}$ with pointwise addition, but scalar multiplication defined by $(af)(x) = f(ax)$.

m. The set V of all 2×2 matrices whose entries sum to 0; operations of $\mathbf{M}_{22}$.

n. The set V of all 2×2 matrices with the addition of $\mathbf{M}_{22}$ but scalar multiplication $*$ defined by $a * X = aX^T$.

Exercise 6.1.3 Let V be the set of positive real numbers with vector addition being ordinary multiplication, and scalar multiplication being $a\cdot v = v^a$. Show that V is a vector space.

Exercise 6.1.4 If V is the set of ordered pairs $(x,\ y)$ of real numbers, show that it is a vector space with addition $(x,\ y) + (x_1,\ y_1) = (x + x_1,\ y + y_1 + 1)$ and scalar multiplication $a(x,\ y) = (ax,\ ay + a - 1)$. What is the zero vector in V?

Exercise 6.1.5 Find $\mathbf{x}$ and $\mathbf{y}$ (in terms of $\mathbf{u}$ and $\mathbf{v}$) such that:

a. $\begin{aligned} 2\mathbf{x} + \ \ \mathbf{y} &= \mathbf{u} \\ 5\mathbf{x} + 3\mathbf{y} &= \mathbf{v} \end{aligned}$

b. $\begin{aligned} 3\mathbf{x} - 2\mathbf{y} &= \mathbf{u} \\ 4\mathbf{x} - 5\mathbf{y} &= \mathbf{v} \end{aligned}$

Exercise 6.1.6 In each case show that the condition $a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = \mathbf{0}$ in V implies that $a = b = c = 0$.

a. $V = \mathbb{R}^4$; $\mathbf{u} = (2,\ 1,\ 0,\ 2)$, $\mathbf{v} = (1,\ 1,\ -1,\ 0)$, $\mathbf{w} = (0,\ 1,\ 2,\ 1)$

b. $V = \mathbf{M}_{22}$; $\mathbf{u} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

c. $V = \mathbf{P}$; $\mathbf{u} = x^3 + x$, $\mathbf{v} = x^2 + 1$, $\mathbf{w} = x^3 - x^2 + x + 1$

d. $V = \mathbf{F}[0,\ \pi]$; $\mathbf{u} = \sin x$, $\mathbf{v} = \cos x$, $\mathbf{w} = 1$—the constant function

Exercise 6.1.7 Simplify each of the following.

a. $3[2(\mathbf{u} - 2\mathbf{v} - \mathbf{w}) + 3(\mathbf{w} - \mathbf{v})] - 7(\mathbf{u} - 3\mathbf{v} - \mathbf{w})$

b. $4(3\mathbf{u} - \mathbf{v} + \mathbf{w}) - 2[(3\mathbf{u} - 2\mathbf{v}) - 3(\mathbf{v} - \mathbf{w})] + 6(\mathbf{w} - \mathbf{u} - \mathbf{v})$

Exercise 6.1.8 Show that $\mathbf{x} = \mathbf{v}$ is the only solution to the equation $\mathbf{x} + \mathbf{x} = 2\mathbf{v}$ in a vector space V. Cite all axioms used.

Exercise 6.1.9 Show that $-\mathbf{0} = \mathbf{0}$ in any vector space. Cite all axioms used.

Exercise 6.1.10 Show that the zero vector $\mathbf{0}$ is uniquely determined by the property in axiom A4.

Exercise 6.1.11 Given a vector $\mathbf{v}$, show that its negative $-\mathbf{v}$ is uniquely determined by the property in axiom A5.

Exercise 6.1.12

a. Prove (2) of Theorem 6.1.3. [*Hint*: Axiom S2.]

b. Prove that $(-a)\mathbf{v} = -(a\mathbf{v})$ in Theorem 6.1.3 by first computing $(-a)\mathbf{v} + a\mathbf{v}$. Then do it using (4) of Theorem 6.1.3 and axiom S4.

c. Prove that $a(-\mathbf{v}) = -(a\mathbf{v})$ in Theorem 6.1.3 in two ways, as in part (b).

Exercise 6.1.13 Let $\mathbf{v}$, $\mathbf{v}_1$, $\dots$, $\mathbf{v}_n$ denote vectors in a vector space V and let a, a_1, $\dots$, a_n denote numbers. Use induction on n to prove each of the following.

a. $a(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n) = a\mathbf{v}_1 + a\mathbf{v}_2 + \cdots + a\mathbf{v}_n$

b. $(a_1 + a_2 + \cdots + a_n)\mathbf{v} = a_1\mathbf{v} + a_2\mathbf{v} + \cdots + a_n\mathbf{v}$

Exercise 6.1.14 Verify axioms A2—A5 and S2—S5 for the space $\mathbf{F}[a, b]$ of functions on $[a, b]$ (Example 6.1.7).

Exercise 6.1.15 Prove each of the following for vectors $\mathbf{u}$ and $\mathbf{v}$ and scalars a and b.

a. If $a\mathbf{v} = \mathbf{0}$, then $a = 0$ or $\mathbf{v} = \mathbf{0}$.

b. If $a\mathbf{v} = b\mathbf{v}$ and $\mathbf{v} \neq \mathbf{0}$, then $a = b$.

c. If $a\mathbf{v} = a\mathbf{w}$ and $a \neq 0$, then $\mathbf{v} = \mathbf{w}$.

Exercise 6.1.16 By calculating $(1+1)(\mathbf{v}+\mathbf{w})$ in two ways (using axioms S2 and S3), show that axiom A2 follows from the other axioms.

Exercise 6.1.17 Let V be a vector space, and define V^n to be the set of all n-tuples $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ of n vectors $\mathbf{v}_i$, each belonging to V. Define addition and scalar multiplication in V^n as follows:

$$\begin{aligned}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) + (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \\ = (\mathbf{u}_1 + \mathbf{v}_1, \mathbf{u}_2 + \mathbf{v}_2, \dots, \mathbf{u}_n + \mathbf{v}_n) \\ a(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = (a\mathbf{v}_1, a\mathbf{v}_2, \dots, a\mathbf{v}_n)\end{aligned}$$

Show that V^n is a vector space.

Exercise 6.1.18 Let V^n be the vector space of n-tuples from the preceding exercise, written as columns. If A is an $m \times n$ matrix, and X is in V^n, define AX in V^m by matrix multiplication. More precisely, if

$$A = [a_{ij}] \text{ and } X = \left[\begin{array}{c}\mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n\end{array}\right], \text{ let } AX = \left[\begin{array}{c}\mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n\end{array}\right]$$

where $\mathbf{u}_i = a_{i1}\mathbf{v}_1 + a_{i2}\mathbf{v}_2 + \cdots + a_{in}\mathbf{v}_n$ for each i.
Prove that:

a. $B(AX) = (BA)X$

b. $(A + A_1)X = AX + A_1X$

c. $A(X + X_1) = AX + AX_1$

d. $(kA)X = k(AX) = A(kX)$ if k is any number

e. $IX = X$ if I is the $n \times n$ identity matrix

f. Let E be an elementary matrix obtained by performing a row operation on the rows of I_n (see Section 2.5). Show that EX is the column resulting from performing that same row operation on the vectors (call them rows) of X. [*Hint*: Lemma 2.5.1.]

6.2 Subspaces and Spanning Sets

Chapter 5 is essentially about the subspaces of $\mathbb{R}^n$. We now extend this notion.

Definition 6.2 Subspaces of a Vector Space

If V is a vector space, a nonempty subset $U \subseteq V$ is called a ***subspace*** *of V if U is itself a vector space using the addition and scalar multiplication of V.*

Subspaces of $\mathbb{R}^n$ (as defined in Section 5.1) are subspaces in the present sense by Example 6.1.3. Moreover, the defining properties for a subspace of $\mathbb{R}^n$ actually *characterize* subspaces in general.

Theorem 6.2.1: Subspace Test

A subset U of a vector space is a subspace of V if and only if it satisfies the following three conditions:

1. *$\mathbf{0}$ lies in U where $\mathbf{0}$ is the zero vector of V.*
2. *If $\mathbf{u}_1$ and $\mathbf{u}_2$ are in U, then $\mathbf{u}_1 + \mathbf{u}_2$ is also in U.*
3. *If $\mathbf{u}$ is in U, then $a\mathbf{u}$ is also in U for each scalar a.*

Proof. If U is a subspace of V, then (2) and (3) hold by axioms A1 and S1 respectively, applied to the vector space U. Since U is nonempty (it is a vector space), choose $\mathbf{u}$ in U. Then (1) holds because $\mathbf{0} = 0\mathbf{u}$ is in U by (3) and Theorem 6.1.3.

Conversely, if (1), (2), and (3) hold, then axioms A1 and S1 hold because of (2) and (3), and axioms A2, A3, S2, S3, S4, and S5 hold in U because they hold in V. Axiom A4 holds because the zero vector $\mathbf{0}$ of V is actually in U by (1), and so serves as the zero of U. Finally, given $\mathbf{u}$ in U, then its negative $-\mathbf{u}$ in V is again in U by (3) because $-\mathbf{u} = (-1)\mathbf{u}$ (again using Theorem 6.1.3). Hence $-\mathbf{u}$ serves as the negative of $\mathbf{u}$ in U. □

Note that the proof of Theorem 6.2.1 shows that if U is a subspace of V, then U and V share the same zero vector, and that the negative of a vector in the space U is the same as its negative in V.

Example 6.2.1

If V is any vector space, show that $\{\mathbf{0}\}$ and V are subspaces of V.

Solution. $U = V$ clearly satisfies the conditions of the subspace test. As to $U = \{\mathbf{0}\}$, it satisfies the conditions because $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $a\mathbf{0} = \mathbf{0}$ for all a in $\mathbb{R}$.

The vector space $\{\mathbf{0}\}$ is called the **zero subspace** of V.

Example 6.2.2

Let $\mathbf{v}$ be a vector in a vector space V. Show that the set

$$\mathbb{R}\mathbf{v} = \{a\mathbf{v} \mid a \text{ in } \mathbb{R}\}$$

of all scalar multiples of $\mathbf{v}$ is a subspace of V.

Solution. Because $\mathbf{0} = 0\mathbf{v}$, it is clear that $\mathbf{0}$ lies in $\mathbb{R}\mathbf{v}$. Given two vectors $a\mathbf{v}$ and $a_1\mathbf{v}$ in $\mathbb{R}\mathbf{v}$, their sum $a\mathbf{v} + a_1\mathbf{v} = (a + a_1)\mathbf{v}$ is also a scalar multiple of $\mathbf{v}$ and so lies in $\mathbb{R}\mathbf{v}$. Hence $\mathbb{R}\mathbf{v}$ is closed under addition. Finally, given $a\mathbf{v}$, $r(a\mathbf{v}) = (ra)\mathbf{v}$ lies in $\mathbb{R}\mathbf{v}$ for all $r \in \mathbb{R}$, so $\mathbb{R}\mathbf{v}$ is closed under scalar multiplication. Hence the subspace test applies.

In particular, given $\mathbf{d} \neq \mathbf{0}$ in $\mathbb{R}^3$, $\mathbb{R}\mathbf{d}$ is the line through the origin with direction vector $\mathbf{d}$.

The space $\mathbb{R}\mathbf{v}$ in Example 6.2.2 is described by giving the *form* of each vector in $\mathbb{R}\mathbf{v}$. The next example describes a subset U of the space $\mathbf{M}_{nn}$ by giving a *condition* that each matrix of U must satisfy.

Example 6.2.3

Let A be a fixed matrix in $\mathbf{M}_{nn}$. Show that $U = \{X \text{ in } \mathbf{M}_{nn} \mid AX = XA\}$ is a subspace of $\mathbf{M}_{nn}$.

Solution. If 0 is the $n \times n$ zero matrix, then $A0 = 0A$, so 0 satisfies the condition for membership in U. Next suppose that X and X_1 lie in U so that $AX = XA$ and $AX_1 = X_1A$. Then

$$A(X + X_1) = AX + AX_1 = XA + X_1A + (X + X_1)A$$
$$A(aX) = a(AX) = a(XA) = (aX)A$$

for all a in $\mathbb{R}$, so both $X + X_1$ and aX lie in U. Hence U is a subspace of $\mathbf{M}_{nn}$.

Suppose $p(x)$ is a polynomial and a is a number. Then the number $p(a)$ obtained by replacing x by a in the expression for $p(x)$ is called the **evaluation** of $p(x)$ at a. For example, if $p(x) = 5 - 6x + 2x^2$, then the evaluation of $p(x)$ at $a = 2$ is $p(2) = 5 - 12 + 8 = 1$. If $p(a) = 0$, the number a is called a **root** of $p(x)$.

Example 6.2.4

Consider the set U of all polynomials in $\mathbf{P}$ that have 3 as a root:

$$U = \{p(x) \in \mathbf{P} \mid p(3) = 0\}$$

Show that U is a subspace of $\mathbf{P}$.

Solution. Clearly, the zero polynomial lies in U. Now let $p(x)$ and $q(x)$ lie in U so $p(3) = 0$ and $q(3) = 0$. We have $(p+q)(x) = p(x) + q(x)$ for all x, so $(p+q)(3) = p(3) + q(3) = 0 + 0 = 0$, and U is closed under addition. The verification that U is closed under scalar multiplication is similar.

Recall that the space $\mathbf{P}_n$ consists of all polynomials of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0, a_1, a_2, \dots, a_n$ are real numbers, and so is closed under the addition and scalar multiplication in $\mathbf{P}$. Moreover, the zero polynomial is included in $\mathbf{P}_n$. Thus the subspace test gives Example 6.2.5.

Example 6.2.5

$\mathbf{P}_n$ is a subspace of $\mathbf{P}$ for each $n \geq 0$.

The next example involves the notion of the derivative f' of a function f. (If the reader is not familiar with calculus, this example may be omitted.) A function f defined on the interval $[a, b]$ is called **differentiable** if the derivative $f'(r)$ exists at every r in $[a, b]$.

Example 6.2.6

Show that the subset $\mathbf{D}[a, b]$ of all **differentiable functions** on $[a, b]$ is a subspace of the vector space $\mathbf{F}[a, b]$ of all functions on $[a, b]$.

Solution. The derivative of any constant function is the constant function 0; in particular, 0 itself is differentiable and so lies in $\mathbf{D}[a, b]$. If f and g both lie in $\mathbf{D}[a, b]$ (so that f' and g' exist), then it is a theorem of calculus that $f+g$ and rf are both differentiable for any $r \in \mathbb{R}$. In fact, $(f+g)' = f'+g'$ and $(rf)' = rf'$, so both lie in $\mathbf{D}[a, b]$. This shows that $\mathbf{D}[a, b]$ is a subspace of $\mathbf{F}[a, b]$.

Linear Combinations and Spanning Sets

Definition 6.3 Linear Combinations and Spanning

*Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a set of vectors in a vector space V. As in $\mathbb{R}^n$, a vector $\mathbf{v}$ is called a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ if it can be expressed in the form*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

*where $a_1, a_2, \dots, a_n$ are scalars, called the **coefficients** of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. The set of all linear combinations of these vectors is called their **span**, and is denoted by*

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \mid a_i \text{ in } \mathbb{R}\}$$

If it happens that $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, these vectors are called a **spanning set** for V. For example, the span of two vectors $\mathbf{v}$ and $\mathbf{w}$ is the set

$$\operatorname{span}\{\mathbf{v}, \mathbf{w}\} = \{s\mathbf{v} + t\mathbf{w} \mid s \text{ and } t \text{ in } \mathbb{R}\}$$

of all sums of scalar multiples of these vectors.

Example 6.2.7

Consider the vectors $p_1 = 1+x+4x^2$ and $p_2 = 1+5x+x^2$ in $\mathbf{P}_2$. Determine whether p_1 and p_2 lie in $\operatorname{span}\{1+2x-x^2, 3+5x+2x^2\}$.

Solution. For p_1, we want to determine if s and t exist such that

$$p_1 = s(1+2x-x^2) + t(3+5x+2x^2)$$

Equating coefficients of powers of x (where $x^0 = 1$) gives

$$1 = s+3t, \qquad 1 = 2s+5t, \qquad \text{and} \qquad 4 = -s+2t$$

These equations have the solution $s = -2$ and $t = 1$, so p_1 is indeed in $\operatorname{span}\{1+2x-x^2, 3+5x+2x^2\}$.

Turning to $p_2 = 1 + 5x + x^2$, we are looking for s and t such that

$$p_2 = s(1 + 2x - x^2) + t(3 + 5x + 2x^2)$$

Again equating coefficients of powers of x gives equations $1 = s + 3t$, $5 = 2s + 5t$, and $1 = -s + 2t$. But in this case there is no solution, so p_2 is *not* in $\text{span}\,\{1 + 2x - x^2,\ 3 + 5x + 2x^2\}$.

We saw in Example 5.1.6 that $\mathbb{R}^m = \text{span}\,\{\mathbf{e}_1,\ \mathbf{e}_2,\ \dots,\ \mathbf{e}_m\}$ where the vectors $\mathbf{e}_1,\ \mathbf{e}_2,\ \dots,\ \mathbf{e}_m$ are the columns of the $m \times m$ identity matrix. Of course $\mathbb{R}^m = \mathbf{M}_{m1}$ is the set of all $m \times 1$ matrices, and there is an analogous spanning set for each space $\mathbf{M}_{mn}$. For example, each 2×2 matrix has the form

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] = a \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \right] + b \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right] + c \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array} \right] + d \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right]$$

so

$$\mathbf{M}_{22} = \text{span}\left\{ \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right] \right\}$$

Similarly, we obtain

Example 6.2.8

$\mathbf{M}_{mn}$ is the span of the set of all $m \times n$ matrices with exactly one entry equal to 1, and all other entries zero.

The fact that every polynomial in $\mathbf{P}_n$ has the form $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ where each a_i is in $\mathbb{R}$ shows that

Example 6.2.9

$\mathbf{P}_n = \text{span}\,\{1,\ x,\ x^2,\ \dots,\ x^n\}$.

In Example 6.2.2 we saw that $\text{span}\,\{\mathbf{v}\} = \{a\mathbf{v} \mid a \text{ in } \mathbb{R}\} = \mathbb{R}\mathbf{v}$ is a subspace for any vector $\mathbf{v}$ in a vector space V. More generally, the span of *any* set of vectors is a subspace. In fact, the proof of Theorem 5.1.1 goes through to prove:

Theorem 6.2.2

Let $U = \text{span}\,\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n\}$ *in a vector space* V*. Then:*

1. U *is a subspace of* V *containing each of* $\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n$.
2. U *is the "smallest" subspace containing these vectors in the sense that any subspace that contains each of* $\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n$ *must contain* U.

Here is how condition 2 in Theorem 6.2.2 is used. Given vectors $\mathbf{v}_1,\ \dots,\ \mathbf{v}_k$ in a vector space V and a subspace $U \subseteq V$, then:

$$\text{span}\,\{\mathbf{v}_1,\ \dots,\ \mathbf{v}_n\} \subseteq U \Leftrightarrow \text{ each } \mathbf{v}_i \in U$$

The following examples illustrate this.

Example 6.2.10

Show that $\mathbf{P}_3 = \text{span}\{x^2 + x^3,\ x,\ 2x^2 + 1,\ 3\}$.

Solution. Write $U = \text{span}\{x^2 + x^3,\ x,\ 2x^2 + 1,\ 3\}$. Then $U \subseteq \mathbf{P}_3$, and we use the fact that $\mathbf{P}_3 = \text{span}\{1,\ x,\ x^2,\ x^3\}$ to show that $\mathbf{P}_3 \subseteq U$. In fact, x and $1 = \frac{1}{3} \cdot 3$ clearly lie in U. But then successively,

$$x^2 = \frac{1}{2}[(2x^2 + 1) - 1] \quad \text{and} \quad x^3 = (x^2 + x^3) - x^2$$

also lie in U. Hence $\mathbf{P}_3 \subseteq U$ by Theorem 6.2.2.

Example 6.2.11

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in a vector space V. Show that

$$\text{span}\{\mathbf{u},\ \mathbf{v}\} = \text{span}\{\mathbf{u} + 2\mathbf{v},\ \mathbf{u} - \mathbf{v}\}$$

Solution. We have $\text{span}\{\mathbf{u} + 2\mathbf{v},\ \mathbf{u} - \mathbf{v}\} \subseteq \text{span}\{\mathbf{u},\ \mathbf{v}\}$ by Theorem 6.2.2 because both $\mathbf{u} + 2\mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ lie in $\text{span}\{\mathbf{u},\ \mathbf{v}\}$. On the other hand,

$$\mathbf{u} = \frac{1}{3}(\mathbf{u} + 2\mathbf{v}) + \frac{2}{3}(\mathbf{u} - \mathbf{v}) \quad \text{and} \quad \mathbf{v} = \frac{1}{3}(\mathbf{u} + 2\mathbf{v}) - \frac{1}{3}(\mathbf{u} - \mathbf{v})$$

so $\text{span}\{\mathbf{u},\ \mathbf{v}\} \subseteq \text{span}\{\mathbf{u} + 2\mathbf{v},\ \mathbf{u} - \mathbf{v}\}$, again by Theorem 6.2.2.

Exercises for 6.2

Exercise 6.2.1 Which of the following are subspaces of $\mathbf{P}_3$? Support your answer.

a. $U = \{f(x) \mid f(x) \in \mathbf{P}_3,\ f(2) = 1\}$

b. $U = \{xg(x) \mid g(x) \in \mathbf{P}_2\}$

c. $U = \{xg(x) \mid g(x) \in \mathbf{P}_3\}$

d. $U = \{xg(x) + (1 - x)h(x) \mid g(x) \text{ and } h(x) \in \mathbf{P}_2\}$

e. $U =$ The set of all polynomials in $\mathbf{P}_3$ with constant term 0

f. $U = \{f(x) \mid f(x) \in \mathbf{P}_3,\ \deg f(x) = 3\}$

Exercise 6.2.2 Which of the following are subspaces of $\mathbf{M}_{22}$? Support your answer.

a. $U = \left\{ \left[\begin{array}{cc} a & b \\ 0 & c \end{array} \right] \middle|\ a,\ b, \text{ and } c \text{ in } \mathbb{R} \right\}$

b. $U = \left\{ \left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \middle|\ a + b = c + d;\ a,\ b,\ c,\ d \text{ in } \mathbb{R} \right\}$

c. $U = \{A \mid A \in \mathbf{M}_{22},\ A = A^T\}$

d. $U = \{A \mid A \in \mathbf{M}_{22},\ AB = 0\}$, B a fixed 2×2 matrix

e. $U = \{A \mid A \in \mathbf{M}_{22},\ A^2 = A\}$

f. $U = \{A \mid A \in \mathbf{M}_{22},\ A \text{ is not invertible}\}$

g. $U = \{A \mid A \in \mathbf{M}_{22},\ BAC = CAB\}$, B and C fixed 2×2 matrices

Exercise 6.2.3 Which of the following are subspaces of $\mathbf{F}[0,\ 1]$? Support your answer.

a. $U = \{f \mid f(0) = 0\}$

b. $U = \{f \mid f(0) = 1\}$

c. $U = \{f \mid f(0) = f(1)\}$

d. $U = \{f \mid f(x) \geq 0$ for all x in $[0,\ 1]\}$

e. $U = \{f \mid f(x) = f(y)$ for all x and y in $[0,\ 1]\}$

f. $U = \{f \mid f(x+y) = f(x) + f(y)$ for all x and y in $[0,\ 1]\}$

g. $U = \{f \mid f$ is integrable and $\int_0^1 f(x)dx = 0\}$

Exercise 6.2.4 Let A be an $m \times n$ matrix. For which columns $\mathbf{b}$ in $\mathbb{R}^m$ is $U = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ A\mathbf{x} = \mathbf{b}\}$ a subspace of $\mathbb{R}^n$? Support your answer.

Exercise 6.2.5 Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$ (written as a column), and define $U = \{A\mathbf{x} \mid A \in \mathbf{M}_{mn}\}$.

a. Show that U is a subspace of $\mathbb{R}^m$.

b. Show that $U = \mathbb{R}^m$ if $\mathbf{x} \neq \mathbf{0}$.

Exercise 6.2.6 Write each of the following as a linear combination of $x+1$, x^2+x, and x^2+2.

a. x^2+3x+2

b. $2x^2-3x+1$

c. x^2+1

d. x

Exercise 6.2.7 Determine whether $\mathbf{v}$ lies in $\operatorname{span}\{\mathbf{u},\ \mathbf{w}\}$ in each case.

a. $\mathbf{v} = 3x^2-2x-1$; $\mathbf{u} = x^2+1$, $\mathbf{w} = x+2$

b. $\mathbf{v} = x$; $\mathbf{u} = x^2+1$, $\mathbf{w} = x+2$

c. $\mathbf{v} = \left[\begin{array}{rr} 1 & 3 \\ -1 & 1 \end{array}\right]$; $\mathbf{u} = \left[\begin{array}{rr} 1 & -1 \\ 2 & 1 \end{array}\right]$, $\mathbf{w} = \left[\begin{array}{rr} 2 & 1 \\ 1 & 0 \end{array}\right]$

d. $\mathbf{v} = \left[\begin{array}{rr} 1 & -4 \\ 5 & 3 \end{array}\right]$; $\mathbf{u} = \left[\begin{array}{rr} 1 & -1 \\ 2 & 1 \end{array}\right]$, $\mathbf{w} = \left[\begin{array}{rr} 2 & 1 \\ 1 & 0 \end{array}\right]$

Exercise 6.2.8 Which of the following functions lie in $\operatorname{span}\{\cos^2 x,\ \sin^2 x\}$? (Work in $\mathbf{F}[0,\ \pi]$.)

a. $\cos 2x$

b. 1

c. x^2

d. $1+x^2$

Exercise 6.2.9

a. Show that $\mathbb{R}^3$ is spanned by $\{(1,\ 0,\ 1),\ (1,\ 1,\ 0),\ (0,\ 1,\ 1)\}$.

b. Show that $\mathbf{P}_2$ is spanned by $\{1+2x^2,\ 3x,\ 1+x\}$.

c. Show that $\mathbf{M}_{22}$ is spanned by
$\left\{ \left[\begin{array}{rr} 1 & 0 \\ 0 & 0 \end{array}\right], \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right], \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \end{array}\right], \left[\begin{array}{rr} 1 & 1 \\ 0 & 1 \end{array}\right] \right\}$.

Exercise 6.2.10 If X and Y are two sets of vectors in a vector space V, and if $X \subseteq Y$, show that $\operatorname{span} X \subseteq \operatorname{span} Y$.

Exercise 6.2.11 Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ denote vectors in a vector space V. Show that:

a. $\operatorname{span}\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\} = \operatorname{span}\{\mathbf{u}+\mathbf{v},\ \mathbf{u}+\mathbf{w},\ \mathbf{v}+\mathbf{w}\}$

b. $\operatorname{span}\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\} = \operatorname{span}\{\mathbf{u}-\mathbf{v},\ \mathbf{u}+\mathbf{w},\ \mathbf{w}\}$

Exercise 6.2.12 Show that

$$\operatorname{span}\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n,\ \mathbf{0}\} = \operatorname{span}\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n\}$$

holds for any set of vectors $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n\}$.

Exercise 6.2.13 If X and Y are nonempty subsets of a vector space V such that $\operatorname{span} X = \operatorname{span} Y = V$, must there be a vector common to both X and Y? Justify your answer.

Exercise 6.2.14 Is it possible that $\{(1,\ 2,\ 0),\ (1,\ 1,\ 1)\}$ can span the subspace $U = \{(a,\ b,\ 0) \mid a$ and b in $\mathbb{R}\}$?

Exercise 6.2.15 Describe $\operatorname{span}\{\mathbf{0}\}$.

Exercise 6.2.16 Let $\mathbf{v}$ denote any vector in a vector space V. Show that $\operatorname{span}\{\mathbf{v}\} = \operatorname{span}\{a\mathbf{v}\}$ for any $a \neq 0$.

Exercise 6.2.17 Determine all subspaces of $\mathbb{R}\mathbf{v}$ where $\mathbf{v} \neq \mathbf{0}$ in some vector space V.

Exercise 6.2.18 Suppose $V = \operatorname{span}\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n\}$. If $\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ where the a_i are in $\mathbb{R}$ and $a_1 \neq 0$, show that $V = \operatorname{span}\{\mathbf{u},\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_n\}$.

Exercise 6.2.19 If $\mathbf{M}_{nn} = \operatorname{span}\{A_1,\ A_2,\ \dots,\ A_k\}$, show that $\mathbf{M}_{nn} = \operatorname{span}\{A_1^T,\ A_2^T,\ \dots,\ A_k^T\}$.

Exercise 6.2.20 If $\mathbf{P}_n = \text{span}\,\{p_1(x),\ p_2(x),\ \ldots,\ p_k(x)\}$ and a is in $\mathbb{R}$, show that $p_i(a) \neq 0$ for some i.

Exercise 6.2.21 Let U be a subspace of a vector space V.

a. If $a\mathbf{u}$ is in U where $a \neq 0$, show that $\mathbf{u}$ is in U.

b. If $\mathbf{u}$ and $\mathbf{u}+\mathbf{v}$ are in U, show that $\mathbf{v}$ is in U.

Exercise 6.2.22 Let U be a nonempty subset of a vector space V. Show that U is a subspace of V if and only if $\mathbf{u}_1 + a\mathbf{u}_2$ lies in U for all $\mathbf{u}_1$ and $\mathbf{u}_2$ in U and all a in $\mathbb{R}$.

Exercise 6.2.23 Let $U = \{p(x) \text{ in } \mathbf{P} \mid p(3) = 0\}$ be the set in Example 6.2.4. Use the factor theorem (see Section 6.5) to show that U consists of multiples of $x-3$; that is, show that $U = \{(x-3)q(x) \mid q(x) \in \mathbf{P}\}$. Use this to show that U is a subspace of $\mathbf{P}$.

Exercise 6.2.24 Let $A_1,\ A_2,\ \ldots,\ A_m$ denote $n \times n$ matrices. If $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$ and $A_1\mathbf{y} = A_2\mathbf{y} = \cdots = A_m\mathbf{y} = \mathbf{0}$, show that $\{A_1,\ A_2,\ \ldots,\ A_m\}$ cannot span $\mathbf{M}_{nn}$.

Exercise 6.2.25 Let $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_n\}$ and $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots,\ \mathbf{u}_n\}$ be sets of vectors in a vector space, and let

$$X = \left[\begin{array}{c} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{array}\right] \quad Y = \left[\begin{array}{c} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{array}\right]$$

as in Exercise 6.1.18.

a. Show that $\text{span}\,\{\mathbf{v}_1,\ \ldots,\ \mathbf{v}_n\} \subseteq \text{span}\,\{\mathbf{u}_1,\ \ldots,\ \mathbf{u}_n\}$ if and only if $AY = X$ for some $n \times n$ matrix A.

b. If $X = AY$ where A is invertible, show that $\text{span}\,\{\mathbf{v}_1,\ \ldots,\ \mathbf{v}_n\} = \text{span}\,\{\mathbf{u}_1,\ \ldots,\ \mathbf{u}_n\}$.

Exercise 6.2.26 If U and W are subspaces of a vector space V, let $U \cup W = \{\mathbf{v} \mid \mathbf{v} \text{ is in } U \text{ or } \mathbf{v} \text{ is in } W\}$. Show that $U \cup W$ is a subspace if and only if $U \subseteq W$ or $W \subseteq U$.

Exercise 6.2.27 Show that $\mathbf{P}$ cannot be spanned by a finite set of polynomials.

6.3 Linear Independence and Dimension

Definition 6.4 Linear Independence and Dependence

*As in $\mathbb{R}^n$, a set of vectors $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_n\}$ in a vector space V is called **linearly independent** (or simply **independent**) if it satisfies the following condition:*

$$\text{If} \quad s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n = \mathbf{0}, \qquad \text{then} \quad s_1 = s_2 = \cdots = s_n = 0.$$

*A set of vectors that is not linearly independent is said to be **linearly dependent** (or simply **dependent**).*

The **trivial linear combination** of the vectors $\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_n$ is the one with every coefficient zero:

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n$$

This is obviously one way of expressing $\mathbf{0}$ as a linear combination of the vectors $\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_n$, and they are linearly independent when it is the *only* way.

Example 6.3.1

Show that $\{1+x,\ 3x+x^2,\ 2+x-x^2\}$ is independent in $\mathbf{P}_2$.

Solution. Suppose a linear combination of these polynomials vanishes.

$$s_1(1+x)+s_2(3x+x^2)+s_3(2+x-x^2)=0$$

Equating the coefficients of 1, x, and x^2 gives a set of linear equations.

$$\begin{array}{rcrcrcl} s_1 & + & & + & 2s_3 & = & 0 \\ s_1 & + & 3s_2 & + & s_3 & = & 0 \\ & & s_2 & - & s_3 & = & 0 \end{array}$$

The only solution is $s_1=s_2=s_3=0$.

Example 6.3.2

Show that $\{\sin x,\ \cos x\}$ is independent in the vector space $\mathbf{F}[0,\ 2\pi]$ of functions defined on the interval $[0,\ 2\pi]$.

Solution. Suppose that a linear combination of these functions vanishes.

$$s_1(\sin x)+s_2(\cos x)=0$$

This must hold for *all* values of x in $[0,\ 2\pi]$ (by the definition of equality in $\mathbf{F}[0,\ 2\pi]$). Taking $x=0$ yields $s_2=0$ (because $\sin 0=0$ and $\cos 0=1$). Similarly, $s_1=0$ follows from taking $x=\frac{\pi}{2}$ (because $\sin\frac{\pi}{2}=1$ and $\cos\frac{\pi}{2}=0$).

Example 6.3.3

Suppose that $\{\mathbf{u},\ \mathbf{v}\}$ is an independent set in a vector space V. Show that $\{\mathbf{u}+2\mathbf{v},\ \mathbf{u}-3\mathbf{v}\}$ is also independent.

Solution. Suppose a linear combination of $\mathbf{u}+2\mathbf{v}$ and $\mathbf{u}-3\mathbf{v}$ vanishes:

$$s(\mathbf{u}+2\mathbf{v})+t(\mathbf{u}-3\mathbf{v})=\mathbf{0}$$

We must deduce that $s=t=0$. Collecting terms involving $\mathbf{u}$ and $\mathbf{v}$ gives

$$(s+t)\mathbf{u}+(2s-3t)\mathbf{v}=\mathbf{0}$$

Because $\{\mathbf{u},\ \mathbf{v}\}$ is independent, this yields linear equations $s+t=0$ and $2s-3t=0$. The only solution is $s=t=0$.

Example 6.3.4

Show that any set of polynomials of distinct degrees is independent.

Solution. Let $p_1, p_2, \dots, p_m$ be polynomials where $\deg(p_i) = d_i$. By relabelling if necessary, we may assume that $d_1 > d_2 > \cdots > d_m$. Suppose that a linear combination vanishes:

$$t_1p_1 + t_2p_2 + \cdots + t_mp_m = 0$$

where each t_i is in $\mathbb{R}$. As $\deg(p_1) = d_1$, let ax^{d_1} be the term in p_1 of highest degree, where $a \neq 0$. Since $d_1 > d_2 > \cdots > d_m$, it follows that $t_1ax^{d_1}$ is the only term of degree d_1 in the linear combination $t_1p_1 + t_2p_2 + \cdots + t_mp_m = 0$. This means that $t_1ax^{d_1} = 0$, whence $t_1a = 0$, hence $t_1 = 0$ (because $a \neq 0$). But then $t_2p_2 + \cdots + t_mp_m = 0$ so we can repeat the argument to show that $t_2 = 0$. Continuing, we obtain $t_i = 0$ for each i, as desired.

Example 6.3.5

Suppose that A is an $n \times n$ matrix such that $A^k = 0$ but $A^{k-1} \neq 0$. Show that $B = \{I, A, A^2, \dots, A^{k-1}\}$ is independent in $\mathbf{M}_{nn}$.

Solution. Suppose $r_0I + r_1A + r_2A^2 + \cdots + r_{k-1}A^{k-1} = 0$. Multiply by A^{k-1}:

$$r_0A^{k-1} + r_1A^k + r_2A^{k+1} + \cdots + r_{k-1}A^{2k-2} = 0$$

Since $A^k = 0$, all the higher powers are zero, so this becomes $r_0A^{k-1} = 0$. But $A^{k-1} \neq 0$, so $r_0 = 0$, and we have $r_1A^1 + r_2A^2 + \cdots + r_{k-1}A^{k-1} = 0$. Now multiply by A^{k-2} to conclude that $r_1 = 0$. Continuing, we obtain $r_i = 0$ for each i, so B is independent.

The next example collects several useful properties of independence for reference.

Example 6.3.6

Let V denote a vector space.

1. If $\mathbf{v} \neq \mathbf{0}$ in V, then $\{\mathbf{v}\}$ is an independent set.
2. No independent set of vectors in V can contain the zero vector.

Solution.

1. Let $t\mathbf{v} = \mathbf{0}$, t in $\mathbb{R}$. If $t \neq 0$, then $\mathbf{v} = 1\mathbf{v} = \frac{1}{t}(t\mathbf{v}) = \frac{1}{t}\mathbf{0} = \mathbf{0}$, contrary to assumption. So $t = 0$.
2. If $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is independent and (say) $\mathbf{v}_2 = \mathbf{0}$, then $0\mathbf{v}_1 + 1\mathbf{v}_2 + \cdots + 0\mathbf{v}_k = \mathbf{0}$ is a nontrivial linear combination that vanishes, contrary to the independence of $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$.

A set of vectors is independent if $\mathbf{0}$ is a linear combination in a unique way. The following theorem shows that *every* linear combination of these vectors has uniquely determined coefficients, and so extends Theorem 5.2.1.

Theorem 6.3.1

Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a linearly independent set of vectors in a vector space V. If a vector $\mathbf{v}$ has two (ostensibly different) representations

$$\begin{aligned}\mathbf{v} &= s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n \\ \mathbf{v} &= t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n\end{aligned}$$

as linear combinations of these vectors, then $s_1 = t_1, s_2 = t_2, \dots, s_n = t_n$. In other words, every vector in V can be written in a unique way as a linear combination of the $\mathbf{v}_i$.

Proof. Subtracting the equations given in the theorem gives

$$(s_1 - t_1)\mathbf{v}_1 + (s_2 - t_2)\mathbf{v}_2 + \cdots + (s_n - t_n)\mathbf{v}_n = \mathbf{0}$$

The independence of $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ gives $s_i - t_i = 0$ for each i, as required. □

The following theorem extends (and proves) Theorem 5.2.4, and is one of the most useful results in linear algebra.

Theorem 6.3.2: Fundamental Theorem

Suppose a vector space V can be spanned by n vectors. If any set of m vectors in V is linearly independent, then $m \leq n$.

Proof. Let $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, and suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is an independent set in V. Then $\mathbf{u}_1 = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ where each a_i is in $\mathbb{R}$. As $\mathbf{u}_1 \neq \mathbf{0}$ (Example 6.3.6), not all of the a_i are zero, say $a_1 \neq 0$ (after relabelling the $\mathbf{v}_i$). Then $V = \text{span}\{\mathbf{u}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n\}$ as the reader can verify. Hence, write $\mathbf{u}_2 = b_1\mathbf{u}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n$. Then some $c_i \neq 0$ because $\{\mathbf{u}_1, \mathbf{u}_2\}$ is independent; so, as before, $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_3, \dots, \mathbf{v}_n\}$, again after possible relabelling of the $\mathbf{v}_i$. If $m > n$, this procedure continues until all the vectors $\mathbf{v}_i$ are replaced by the vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$. In particular, $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$. But then $\mathbf{u}_{n+1}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ contrary to the independence of the $\mathbf{u}_i$. Hence, the assumption $m > n$ cannot be valid, so $m \leq n$ and the theorem is proved. □

If $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, and if $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is an independent set in V, the above proof shows not only that $m \leq n$ but also that m of the (spanning) vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ can be replaced by the (independent) vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ and the resulting set will still span V. In this form the result is called the **Steinitz Exchange Lemma**.

Definition 6.5 Basis of a Vector Space

As in $\mathbb{R}^n$, a set $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ of vectors in a vector space V is called a ***basis*** *of V if it satisfies the following two conditions:*

1. $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ *is linearly independent*
2. $V = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$

Thus if a set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis, then *every* vector in V can be written as a linear combination of these vectors in a *unique* way (Theorem 6.3.1). But even more is true: Any two (finite) bases of V contain the same number of vectors.

Theorem 6.3.3: Invariance Theorem

Let $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ be two bases of a vector space V. Then $n = m$.

Proof. Because $V = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ is independent, it follows from Theorem 6.3.2 that $m \leq n$. Similarly $n \leq m$, so $n = m$, as asserted. □

Theorem 6.3.3 guarantees that no matter which basis of V is chosen it contains the same number of vectors as any other basis. Hence there is no ambiguity about the following definition.

Definition 6.6 Dimension of a Vector Space

If $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis of the nonzero vector space V, the number n of vectors in the basis is called the ***dimension*** *of V, and we write*

$$\dim V = n$$

The zero vector space $\{\mathbf{0}\}$ is defined to have dimension 0:

$$\dim\{\mathbf{0}\} = 0$$

In our discussion to this point we have always assumed that a basis is nonempty and hence that the dimension of the space is at least 1. However, the zero space $\{\mathbf{0}\}$ has *no* basis (by Example 6.3.6) so our insistence that $\dim\{\mathbf{0}\} = 0$ amounts to saying that the *empty* set of vectors is a basis of $\{\mathbf{0}\}$. Thus the statement that "the dimension of a vector space is the number of vectors in any basis" holds even for the zero space.

We saw in Example 5.2.9 that $\dim(\mathbb{R}^n) = n$ and, if $\mathbf{e}_j$ denotes column j of I_n, that $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis (called the standard basis). In Example 6.3.7 below, similar considerations apply to the space $\mathbf{M}_{mn}$ of all $m \times n$ matrices; the verifications are left to the reader.

Example 6.3.7

The space $\mathbf{M}_{mn}$ has dimension mn, and one basis consists of all $m \times n$ matrices with exactly one entry equal to 1 and all other entries equal to 0. We call this the **standard basis** of $\mathbf{M}_{mn}$.

Example 6.3.8

Show that $\dim \mathbf{P}_n = n+1$ and that $\{1, x, x^2, \dots, x^n\}$ is a basis, called the **standard basis** of $\mathbf{P}_n$.

Solution. Each polynomial $p(x) = a_0 + a_1x + \cdots + a_nx^n$ in $\mathbf{P}_n$ is clearly a linear combination of $1, x, \dots, x^n$, so $\mathbf{P}_n = \text{span}\,\{1, x, \dots, x^n\}$. However, if a linear combination of these vectors vanishes, $a_01 + a_1x + \cdots + a_nx^n = 0$, then $a_0 = a_1 = \cdots = a_n = 0$ because x is an indeterminate. So $\{1, x, \dots, x^n\}$ is linearly independent and hence is a basis containing $n+1$ vectors. Thus, $\dim(\mathbf{P}_n) = n+1$.

Example 6.3.9

If $\mathbf{v} \neq \mathbf{0}$ is any nonzero vector in a vector space V, show that $\text{span}\,\{\mathbf{v}\} = \mathbb{R}\mathbf{v}$ has dimension 1.

Solution. $\{\mathbf{v}\}$ clearly spans $\mathbb{R}\mathbf{v}$, and it is linearly independent by Example 6.3.6. Hence $\{\mathbf{v}\}$ is a basis of $\mathbb{R}\mathbf{v}$, and so $\dim \mathbb{R}\mathbf{v} = 1$.

Example 6.3.10

Let $A = \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right]$ and consider the subspace

$$U = \{X \text{ in } \mathbf{M}_{22} \mid AX = XA\}$$

of $\mathbf{M}_{22}$. Show that $\dim U = 2$ and find a basis of U.

Solution. It was shown in Example 6.2.3 that U is a subspace for any choice of the matrix A. In the present case, if $X = \left[\begin{array}{cc} x & y \\ z & w \end{array} \right]$ is in U, the condition $AX = XA$ gives $z = 0$ and $x = y + w$. Hence each matrix X in U can be written

$$X = \left[\begin{array}{cc} y+w & y \\ 0 & w \end{array} \right] = y \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right] + w \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right]$$

so $U = \text{span}\, B$ where $B = \left\{ \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \right\}$. Moreover, the set B is linearly independent (verify this), so it is a basis of U and $\dim U = 2$.

Example 6.3.11

Show that the set V of all symmetric 2×2 matrices is a vector space, and find the dimension of V.

Solution. A matrix A is symmetric if $A^T = A$. If A and B lie in V, then

$$(A+B)^T = A^T + B^T = A + B \quad \text{and} \quad (kA)^T = kA^T = kA$$

using Theorem 2.1.2. Hence $A+B$ and kA are also symmetric. As the 2×2 zero matrix is also in

V, this shows that V is a vector space (being a subspace of $\mathbf{M}_{22}$). Now a matrix A is symmetric when entries directly across the main diagonal are equal, so each 2×2 symmetric matrix has the form

$$\left[\begin{array}{cc} a & c \\ c & b \end{array}\right] = a\left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right] + b\left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array}\right] + c\left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right]$$

Hence the set $B = \left\{ \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] \right\}$ spans V, and the reader can verify that B is linearly independent. Thus B is a basis of V, so $\dim V = 3$.

It is frequently convenient to alter a basis by multiplying each basis vector by a nonzero scalar. The next example shows that this always produces another basis. The proof is left as Exercise 6.3.22.

Example 6.3.12

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be nonzero vectors in a vector space V. Given nonzero scalars $a_1, a_2, \dots, a_n$, write $D = \{a_1\mathbf{v}_1, a_2\mathbf{v}_2, \dots, a_n\mathbf{v}_n\}$. If B is independent or spans V, the same is true of D. In particular, if B is a basis of V, so also is D.

Exercises for 6.3

Exercise 6.3.1 Show that each of the following sets of vectors is independent.

a. $\{1+x, 1-x, x+x^2\}$ in $\mathbf{P}_2$

b. $\{x^2, x+1, 1-x-x^2\}$ in $\mathbf{P}_2$

c. $\left\{ \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 1 & -1 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 0 & 1 \end{array}\right] \right\}$ in $\mathbf{M}_{22}$

d. $\left\{ \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right] \right\}$ in $\mathbf{M}_{22}$

Exercise 6.3.2 Which of the following subsets of V are independent?

a. $V = \mathbf{P}_2$; $\{x^2+1, x+1, x\}$

b. $V = \mathbf{P}_2$; $\{x^2-x+3, 2x^2+x+5, x^2+5x+1\}$

c. $V = \mathbf{M}_{22}$; $\left\{ \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \right\}$

d. $V = \mathbf{M}_{22}$; $\left\{ \left[\begin{array}{cc} -1 & 0 \\ 0 & -1 \end{array}\right], \left[\begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array}\right], \left[\begin{array}{cc} 0 & -1 \\ -1 & 0 \end{array}\right] \right\}$

e. $V = \mathbf{F}[1, 2]$; $\left\{\frac{1}{x}, \frac{1}{x^2}, \frac{1}{x^3}\right\}$

f. $V = \mathbf{F}[0, 1]$; $\left\{\frac{1}{x^2+x-6}, \frac{1}{x^2-5x+6}, \frac{1}{x^2-9}\right\}$

Exercise 6.3.3 Which of the following are independent in $\mathbf{F}[0, 2\pi]$?

a. $\{\sin^2 x, \cos^2 x\}$

b. $\{1, \sin^2 x, \cos^2 x\}$

c. $\{x, \sin^2 x, \cos^2 x\}$

Exercise 6.3.4 Find all values of a such that the following are independent in $\mathbb{R}^3$.

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

b. $\{(2,\ a,\ 1),\ (1,\ 0,\ 1),\ (0,\ 1,\ 3)\}$

Exercise 6.3.5 Show that the following are bases of the space V indicated.

a. $\{(1,\ 1,\ 0),\ (1,\ 0,\ 1),\ (0,\ 1,\ 1)\};\ V=\mathbb{R}^3$

b. $\{(-1,\ 1,\ 1),\ (1,\ -1,\ 1),\ (1,\ 1,\ -1)\};\ V=\mathbb{R}^3$

c. $\left\{\left[\begin{array}{cc}1&0\\0&1\end{array}\right], \left[\begin{array}{cc}0&1\\1&0\end{array}\right], \left[\begin{array}{cc}1&1\\0&1\end{array}\right], \left[\begin{array}{cc}1&0\\0&0\end{array}\right]\right\};$
$V=\mathbf{M}_{22}$

d. $\{1+x,\ x+x^2,\ x^2+x^3,\ x^3\};\ V=\mathbf{P}_3$

Exercise 6.3.6 Exhibit a basis and calculate the dimension of each of the following subspaces of $\mathbf{P}_2$.

a. $\{a(1+x)+b(x+x^2)\mid a \text{ and } b \text{ in } \mathbb{R}\}$

b. $\{a+b(x+x^2)\mid a \text{ and } b \text{ in } \mathbb{R}\}$

c. $\{p(x)\mid p(1)=0\}$

d. $\{p(x)\mid p(x)=p(-x)\}$

Exercise 6.3.7 Exhibit a basis and calculate the dimension of each of the following subspaces of $\mathbf{M}_{22}$.

a. $\{A\mid A^T=-A\}$

b. $\left\{A \,\middle|\, A\left[\begin{array}{rr}1&1\\-1&0\end{array}\right]=\left[\begin{array}{rr}1&1\\-1&0\end{array}\right]A\right\}$

c. $\left\{A \,\middle|\, A\left[\begin{array}{rr}1&0\\-1&0\end{array}\right]=\left[\begin{array}{rr}0&0\\0&0\end{array}\right]\right\}$

d. $\left\{A \,\middle|\, A\left[\begin{array}{rr}1&1\\-1&0\end{array}\right]=\left[\begin{array}{rr}0&1\\-1&1\end{array}\right]A\right\}$

Exercise 6.3.8 Let $A=\left[\begin{array}{cc}1&1\\0&0\end{array}\right]$ and define $U=\{X\mid X\in\mathbf{M}_{22} \text{ and } AX=X\}$.

a. Find a basis of U containing A.

b. Find a basis of U not containing A.

Exercise 6.3.9 Show that the set $\mathbb{C}$ of all complex numbers is a vector space with the usual operations, and find its dimension.

Exercise 6.3.10

a. Let V denote the set of all 2×2 matrices with equal column sums. Show that V is a subspace of $\mathbf{M}_{22}$, and compute $\dim V$.

b. Repeat part (a) for 3×3 matrices.

c. Repeat part (a) for $n\times n$ matrices.

Exercise 6.3.11

a. Let $V=\{(x^2+x+1)p(x)\mid p(x) \text{ in } \mathbf{P}_2\}$. Show that V is a subspace of $\mathbf{P}_4$ and find $\dim V$. [*Hint*: If $f(x)g(x)=0$ in $\mathbf{P}$, then $f(x)=0$ or $g(x)=0$.]

b. Repeat with $V=\{(x^2-x)p(x)\mid p(x) \text{ in } \mathbf{P}_3\}$, a subset of $\mathbf{P}_5$.

c. Generalize.

Exercise 6.3.12 In each case, either prove the assertion or give an example showing that it is false.

a. Every set of four nonzero polynomials in $\mathbf{P}_3$ is a basis.

b. $\mathbf{P}_2$ has a basis of polynomials $f(x)$ such that $f(0)=0$.

c. $\mathbf{P}_2$ has a basis of polynomials $f(x)$ such that $f(0)=1$.

d. Every basis of $\mathbf{M}_{22}$ contains a noninvertible matrix.

e. No independent subset of $\mathbf{M}_{22}$ contains a matrix A with $A^2=0$.

f. If $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent then, $a\mathbf{u}+b\mathbf{v}+c\mathbf{w}=\mathbf{0}$ for some a, b, c.

g. $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent if $a\mathbf{u}+b\mathbf{v}+c\mathbf{w}=\mathbf{0}$ for some a, b, c.

h. If $\{\mathbf{u},\ \mathbf{v}\}$ is independent, so is $\{\mathbf{u},\ \mathbf{u}+\mathbf{v}\}$.

i. If $\{\mathbf{u},\ \mathbf{v}\}$ is independent, so is $\{\mathbf{u},\ \mathbf{v},\ \mathbf{u}+\mathbf{v}\}$.

j. If $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent, so is $\{\mathbf{u},\ \mathbf{v}\}$.

k. If $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent, so is $\{\mathbf{u}+\mathbf{w},\ \mathbf{v}+\mathbf{w}\}$.

l. If $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent, so is $\{\mathbf{u}+\mathbf{v}+\mathbf{w}\}$.

m. If $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ then $\{\mathbf{u}, \mathbf{v}\}$ is dependent if and only if one is a scalar multiple of the other.

n. If $\dim V = n$, then no set of more than n vectors can be independent.

o. If $\dim V = n$, then no set of fewer than n vectors can span V.

Exercise 6.3.13 Let $A \neq 0$ and $B \neq 0$ be $n \times n$ matrices, and assume that A is symmetric and B is skew-symmetric (that is, $B^T = -B$). Show that $\{A, B\}$ is independent.

Exercise 6.3.14 Show that every set of vectors containing a dependent set is again dependent.

Exercise 6.3.15 Show that every nonempty subset of an independent set of vectors is again independent.

Exercise 6.3.16 Let f and g be functions on $[a, b]$, and assume that $f(a) = 1 = g(b)$ and $f(b) = 0 = g(a)$. Show that $\{f, g\}$ is independent in $\mathbf{F}[a, b]$.

Exercise 6.3.17 Let $\{A_1, A_2, \dots, A_k\}$ be independent in $\mathbf{M}_{mn}$, and suppose that U and V are invertible matrices of size $m \times m$ and $n \times n$, respectively. Show that $\{UA_1V, UA_2V, \dots, UA_kV\}$ is independent.

Exercise 6.3.18 Show that $\{\mathbf{v}, \mathbf{w}\}$ is independent if and only if neither $\mathbf{v}$ nor $\mathbf{w}$ is a scalar multiple of the other.

Exercise 6.3.19 Assume that $\{\mathbf{u}, \mathbf{v}\}$ is independent in a vector space V. Write $\mathbf{u}' = a\mathbf{u} + b\mathbf{v}$ and $\mathbf{v}' = c\mathbf{u} + d\mathbf{v}$, where a, b, c, and d are numbers. Show that $\{\mathbf{u}', \mathbf{v}'\}$ is independent if and only if the matrix $\left[\begin{array}{cc} a & c \\ b & d \end{array} \right]$ is invertible. [*Hint*: Theorem 2.4.5.]

Exercise 6.3.20 If $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is independent and $\mathbf{w}$ is not in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, show that:

a. $\{\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is independent.

b. $\{\mathbf{v}_1 + \mathbf{w}, \mathbf{v}_2 + \mathbf{w}, \dots, \mathbf{v}_k + \mathbf{w}\}$ is independent.

Exercise 6.3.21 If $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is independent, show that $\{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2, \dots, \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_k\}$ is also independent.

Exercise 6.3.22 Prove Example 6.3.12.

Exercise 6.3.23 Let $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ be independent. Which of the following are dependent?

a. $\{\mathbf{u} - \mathbf{v}, \mathbf{v} - \mathbf{w}, \mathbf{w} - \mathbf{u}\}$

b. $\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{u}\}$

c. $\{\mathbf{u} - \mathbf{v}, \mathbf{v} - \mathbf{w}, \mathbf{w} - \mathbf{z}, \mathbf{z} - \mathbf{u}\}$

d. $\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{z}, \mathbf{z} + \mathbf{u}\}$

Exercise 6.3.24 Let U and W be subspaces of V with bases $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ respectively. If U and W have only the zero vector in common, show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{w}_1, \mathbf{w}_2\}$ is independent.

Exercise 6.3.25 Let $\{p, q\}$ be independent polynomials. Show that $\{p, q, pq\}$ is independent if and only if $\deg p \geq 1$ and $\deg q \geq 1$.

Exercise 6.3.26 If z is a complex number, show that $\{z, z^2\}$ is independent if and only if z is not real.

Exercise 6.3.27 Let $B = \{A_1, A_2, \dots, A_n\} \subseteq \mathbf{M}_{mn}$, and write $B' = \{A_1^T, A_2^T, \dots, A_n^T\} \subseteq \mathbf{M}_{nm}$. Show that:

a. B is independent if and only if B' is independent.

b. B spans $\mathbf{M}_{mn}$ if and only if B' spans $\mathbf{M}_{nm}$.

Exercise 6.3.28 If $V = \mathbf{F}[a, b]$ as in Example 6.1.7, show that the set of constant functions is a subspace of dimension 1 (f is **constant** if there is a number c such that $f(x) = c$ for all x).

Exercise 6.3.29

a. If U is an invertible $n \times n$ matrix and $\{A_1, A_2, \dots, A_{mn}\}$ is a basis of $\mathbf{M}_{mn}$, show that $\{A_1U, A_2U, \dots, A_{mn}U\}$ is also a basis.

b. Show that part (a) fails if U is not invertible. [*Hint*: Theorem 2.4.5.]

Exercise 6.3.30 Show that $\{(a, b), (a_1, b_1)\}$ is a basis of $\mathbb{R}^2$ if and only if $\{a + bx, a_1 + b_1x\}$ is a basis of $\mathbf{P}_1$.

Exercise 6.3.31 Find the dimension of the subspace $\text{span}\{1, \sin^2\theta, \cos 2\theta\}$ of $\mathbf{F}[0, 2\pi]$.

Exercise 6.3.32 Show that $\mathbf{F}[0, 1]$ is not finite dimensional.

Exercise 6.3.33 If U and W are subspaces of V, define their intersection $U \cap W$ as follows:

$$U \cap W = \{\mathbf{v} \mid \mathbf{v} \text{ is in both } U \text{ and } W\}$$

a. Show that $U \cap W$ is a subspace contained in U and W.

b. Show that $U \cap W = \{\mathbf{0}\}$ if and only if $\{\mathbf{u},\ \mathbf{w}\}$ is independent for any nonzero vectors $\mathbf{u}$ in U and $\mathbf{w}$ in W.

c. If B and D are bases of U and W, and if $U \cap W = \{\mathbf{0}\}$, show that $B \cup D = \{\mathbf{v} \mid \mathbf{v} \text{ is in } B \text{ or } D\}$ is independent.

Exercise 6.3.34 If U and W are vector spaces, let $V = \{(\mathbf{u},\ \mathbf{w}) \mid \mathbf{u} \text{ in } U \text{ and } \mathbf{w} \text{ in } W\}$.

a. Show that V is a vector space if $(\mathbf{u},\ \mathbf{w}) + (\mathbf{u}_1,\ \mathbf{w}_1) = (\mathbf{u}+\mathbf{u}_1,\ \mathbf{w}+\mathbf{w}_1)$ and $a(\mathbf{u},\ \mathbf{w}) = (a\mathbf{u},\ a\mathbf{w})$.

b. If $\dim U = m$ and $\dim W = n$, show that $\dim V = m+n$.

c. If $V_1,\ \dots,\ V_m$ are vector spaces, let

$$\begin{aligned} V &= V_1 \times \cdots \times V_m \\ &= \{(\mathbf{v}_1,\ \dots,\ \mathbf{v}_m) \mid \mathbf{v}_i \in V_i \text{ for each } i\} \end{aligned}$$

denote the space of n-tuples from the V_i with componentwise operations (see Exercise 6.1.17). If $\dim V_i = n_i$ for each i, show that $\dim V = n_1 + \cdots + n_m$.

Exercise 6.3.35 Let $\mathbf{D}_n$ denote the set of all functions f from the set $\{1,\ 2,\ \dots,\ n\}$ to $\mathbb{R}$.

a. Show that $\mathbf{D}_n$ is a vector space with pointwise addition and scalar multiplication.

b. Show that $\{S_1,\ S_2,\ \dots,\ S_n\}$ is a basis of $\mathbf{D}_n$ where, for each $k = 1,\ 2,\ \dots,\ n$, the function S_k is defined by $S_k(k) = 1$, whereas $S_k(j) = 0$ if $j \neq k$.

Exercise 6.3.36 A polynomial $p(x)$ is called **even** if $p(-x) = p(x)$ and **odd** if $p(-x) = -p(x)$. Let E_n and O_n denote the sets of even and odd polynomials in $\mathbf{P}_n$.

a. Show that E_n is a subspace of $\mathbf{P}_n$ and find $\dim E_n$.

b. Show that O_n is a subspace of $\mathbf{P}_n$ and find $\dim O_n$.

Exercise 6.3.37 Let $\{\mathbf{v}_1,\ \dots,\ \mathbf{v}_n\}$ be independent in a vector space V, and let A be an $n \times n$ matrix. Define $\mathbf{u}_1,\ \dots,\ \mathbf{u}_n$ by

$$\left[\begin{array}{c} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{array}\right] = A \left[\begin{array}{c} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{array}\right]$$

(See Exercise 6.1.18.) Show that $\{\mathbf{u}_1,\ \dots,\ \mathbf{u}_n\}$ is independent if and only if A is invertible.

6.4 Finite Dimensional Spaces

Up to this point, we have had no guarantee that an arbitrary vector space *has* a basis—and hence no guarantee that one can speak *at all* of the dimension of V. However, Theorem 6.4.1 will show that any space that is spanned by a finite set of vectors has a (finite) basis: The proof requires the following basic lemma, of interest in itself, that gives a way to enlarge a given independent set of vectors.

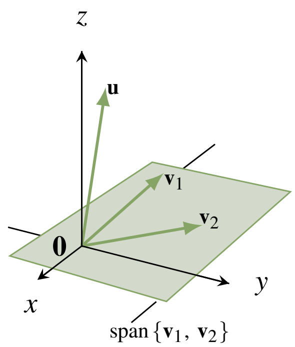
Lemma 6.4.1: Independent Lemma

Let $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_k\}$ be an independent set of vectors in a vector space V. If $\mathbf{u} \in V$ but[5] *$\mathbf{u} \notin \operatorname{span}\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_k\}$, then $\{\mathbf{u},\ \mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_k\}$ is also independent.*

Proof. Let $t\mathbf{u} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$; we must show that all the coefficients are zero. First, $t = 0$ because, otherwise, $\mathbf{u} = -\frac{t_1}{t}\mathbf{v}_1 - \frac{t_2}{t}\mathbf{v}_2 - \cdots - \frac{t_k}{t}\mathbf{v}_k$ is in $\operatorname{span}\{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_k\}$, contrary to our assumption.

[5]If X is a set, we write $a \in X$ to indicate that a is an element of the set X. If a is not an element of X, we write $a \notin X$.

Hence $t = 0$. But then $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$ so the rest of the t_i are zero by the independence of $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$. This is what we wanted. □



Note that the converse of Lemma 6.4.1 is also true: if $\{\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is independent, then $\mathbf{u}$ is not in $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$.

As an illustration, suppose that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is independent in $\mathbb{R}^3$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel, so $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is a plane through the origin (shaded in the diagram). By Lemma 6.4.1, $\mathbf{u}$ is not in this plane if and only if $\{\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2\}$ is independent.

Definition 6.7 Finite Dimensional and Infinite Dimensional Vector Spaces

*A vector space V is called **finite dimensional** if it is spanned by a finite set of vectors. Otherwise, V is called **infinite dimensional**.*

Thus the zero vector space $\{\mathbf{0}\}$ is finite dimensional because $\{\mathbf{0}\}$ is a spanning set.

Lemma 6.4.2

Let V be a finite dimensional vector space. If U is any subspace of V, then any independent subset of U can be enlarged to a finite basis of U.

Proof. Suppose that I is an independent subset of U. If $\operatorname{span} I = U$ then I is already a basis of U. If $\operatorname{span} I \neq U$, choose $\mathbf{u}_1 \in U$ such that $\mathbf{u}_1 \notin \operatorname{span} I$. Hence the set $I \cup \{\mathbf{u}_1\}$ is independent by Lemma 6.4.1. If $\operatorname{span}(I \cup \{\mathbf{u}_1\}) = U$ we are done; otherwise choose $\mathbf{u}_2 \in U$ such that $\mathbf{u}_2 \notin \operatorname{span}(I \cup \{\mathbf{u}_1\})$. Hence $I \cup \{\mathbf{u}_1, \mathbf{u}_2\}$ is independent, and the process continues. We claim that a basis of U will be reached eventually. Indeed, if no basis of U is ever reached, the process creates arbitrarily large independent sets in V. But this is impossible by the fundamental theorem because V is finite dimensional and so is spanned by a finite set of vectors. □

Theorem 6.4.1

Let V be a finite dimensional vector space spanned by m vectors.

1. *V has a finite basis, and $\dim V \leq m$.*
2. *Every independent set of vectors in V can be enlarged to a basis of V by adding vectors from any fixed basis of V.*
3. *If U is a subspace of V, then*
   a. *U is finite dimensional and $\dim U \leq \dim V$.*
   b. *If $\dim U = \dim V$ then $U = V$.*

Proof.

1. If $V = \{\mathbf{0}\}$, then V has an empty basis and $\dim V = 0 \leq m$. Otherwise, let $\mathbf{v} \neq \mathbf{0}$ be a vector in V. Then $\{\mathbf{v}\}$ is independent, so (1) follows from Lemma 6.4.2 with $U = V$.

2. We refine the proof of Lemma 6.4.2. Fix a basis B of V and let I be an independent subset of V. If $\text{span}\, I = V$ then I is already a basis of V. If $\text{span}\, I \neq V$, then B is not contained in I (because B spans V). Hence choose $\mathbf{b}_1 \in B$ such that $\mathbf{b}_1 \notin \text{span}\, I$. Hence the set $I \cup \{\mathbf{b}_1\}$ is independent by Lemma 6.4.1. If $\text{span}\,(I \cup \{\mathbf{b}_1\}) = V$ we are done; otherwise a similar argument shows that $(I \cup \{\mathbf{b}_1,\ \mathbf{b}_2\})$ is independent for some $\mathbf{b}_2 \in B$. Continue this process. As in the proof of Lemma 6.4.2, a basis of V will be reached eventually.

3.
   a. This is clear if $U = \{\mathbf{0}\}$. Otherwise, let $\mathbf{u} \neq \mathbf{0}$ in U. Then $\{\mathbf{u}\}$ can be enlarged to a finite basis B of U by Lemma 6.4.2, proving that U is finite dimensional. But B is independent in V, so $\dim U \leq \dim V$ by the fundamental theorem.

   b. This is clear if $U = \{\mathbf{0}\}$ because V *has* a basis; otherwise, it follows from (2). □

Theorem 6.4.1 shows that a vector space V is finite dimensional if and only if it has a finite basis (possibly empty), and that every subspace of a finite dimensional space is again finite dimensional.

Example 6.4.1

Enlarge the independent set $D = \left\{ \left[\begin{array}{rr} 1 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{rr} 0 & 1 \\ 1 & 1 \end{array} \right], \left[\begin{array}{rr} 1 & 0 \\ 1 & 1 \end{array} \right] \right\}$ to a basis of $\mathbf{M}_{22}$.

Solution. The standard basis of $\mathbf{M}_{22}$ is $\left\{ \left[\begin{array}{rr} 1 & 0 \\ 0 & 0 \end{array} \right], \left[\begin{array}{rr} 0 & 1 \\ 0 & 0 \end{array} \right], \left[\begin{array}{rr} 0 & 0 \\ 1 & 0 \end{array} \right], \left[\begin{array}{rr} 0 & 0 \\ 0 & 1 \end{array} \right] \right\}$, so including one of these in D will produce a basis by Theorem 6.4.1. In fact including *any* of these matrices in D produces an independent set (verify), and hence a basis by Theorem 6.4.4. Of course these vectors are not the only possibilities, for example, including $\left[\begin{array}{rr} 1 & 1 \\ 0 & 1 \end{array} \right]$ works as well.

Example 6.4.2

Find a basis of $\mathbf{P}_3$ containing the independent set $\{1 + x,\ 1 + x^2\}$.

Solution. The standard basis of $\mathbf{P}_3$ is $\{1,\ x,\ x^2,\ x^3\}$, so including two of these vectors will do. If we use 1 and x^3, the result is $\{1,\ 1 + x,\ 1 + x^2,\ x^3\}$. This is independent because the polynomials have distinct degrees (Example 6.3.4), and so is a basis by Theorem 6.4.1. Of course, including $\{1,\ x\}$ or $\{1,\ x^2\}$ would *not* work!

Example 6.4.3

Show that the space $\mathbf{P}$ of all polynomials is infinite dimensional.

Solution. For each $n \geq 1$, $\mathbf{P}$ has a subspace $\mathbf{P}_n$ of dimension $n+1$. Suppose $\mathbf{P}$ is finite dimensional, say $\dim \mathbf{P} = m$. Then $\dim \mathbf{P}_n \leq \dim \mathbf{P}$ by Theorem 6.4.1, that is $n+1 \leq m$. This is impossible since n is arbitrary, so $\mathbf{P}$ must be infinite dimensional.

The next example illustrates how (2) of Theorem 6.4.1 can be used.

Example 6.4.4

If $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k$ are independent columns in $\mathbb{R}^n$, show that they are the first k columns in some invertible $n \times n$ matrix.

Solution. By Theorem 6.4.1, expand $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ to a basis $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k, \mathbf{c}_{k+1}, \dots, \mathbf{c}_n\}$ of $\mathbb{R}^n$. Then the matrix $A = \left[\begin{array}{cccccc} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_k & \mathbf{c}_{k+1} & \dots & \mathbf{c}_n \end{array} \right]$ with this basis as its columns is an $n \times n$ matrix and it is invertible by Theorem 5.2.3.

Theorem 6.4.2

Let U and W be subspaces of the finite dimensional space V.

1. *If $U \subseteq W$, then $\dim U \leq \dim W$.*
2. *If $U \subseteq W$ and $\dim U = \dim W$, then $U = W$.*

Proof. Since W is finite dimensional, (1) follows by taking $V = W$ in part (3) of Theorem 6.4.1. Now assume $\dim U = \dim W = n$, and let B be a basis of U. Then B is an independent set in W. If $U \neq W$, then $\operatorname{span} B \neq W$, so B can be extended to an independent set of $n+1$ vectors in W by Lemma 6.4.1. This contradicts the fundamental theorem (Theorem 6.3.2) because W is spanned by $\dim W = n$ vectors. Hence $U = W$, proving (2). □

Theorem 6.4.2 is very useful. This was illustrated in Example 5.2.13 for $\mathbb{R}^2$ and $\mathbb{R}^3$; here is another example.

Example 6.4.5

If a is a number, let W denote the subspace of all polynomials in $\mathbf{P}_n$ that have a as a root:

$$W = \{p(x) \mid p(x) \in \mathbf{P}_n \text{ and } p(a) = 0\}$$

Show that $\{(x-a), (x-a)^2, \dots, (x-a)^n\}$ is a basis of W.

Solution. Observe first that $(x-a), (x-a)^2, \dots, (x-a)^n$ are members of W, and that they are independent because they have distinct degrees (Example 6.3.4). Write

$$U = \operatorname{span}\{(x-a), (x-a)^2, \dots, (x-a)^n\}$$

Then we have $U \subseteq W \subseteq \mathbf{P}_n$, $\dim U = n$, and $\dim \mathbf{P}_n = n+1$. Hence $n \leq \dim W \leq n+1$ by Theorem 6.4.2. Since $\dim W$ is an integer, we must have $\dim W = n$ or $\dim W = n+1$. But then $W = U$ or $W = \mathbf{P}_n$, again by Theorem 6.4.2. Because $W \neq \mathbf{P}_n$, it follows that $W = U$, as required.

A set of vectors is called **dependent** if it is *not* independent, that is if some nontrivial linear combination vanishes. The next result is a convenient test for dependence.

Lemma 6.4.3: Dependent Lemma

A set $D = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_k\}$ of vectors in a vector space V is dependent if and only if some vector in D is a linear combination of the others.

Proof. Let $\mathbf{v}_2$ (say) be a linear combination of the rest: $\mathbf{v}_2 = s_1\mathbf{v}_1 + s_3\mathbf{v}_3 + \cdots + s_k\mathbf{v}_k$. Then

$$s_1\mathbf{v}_1 + (-1)\mathbf{v}_2 + s_3\mathbf{v}_3 + \cdots + s_k\mathbf{v}_k = \mathbf{0}$$

is a nontrivial linear combination that vanishes, so D is dependent. Conversely, if D is dependent, let $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$ where some coefficient is nonzero. If (say) $t_2 \neq 0$, then $\mathbf{v}_2 = -\frac{t_1}{t_2}\mathbf{v}_1 - \frac{t_3}{t_2}\mathbf{v}_3 - \cdots - \frac{t_k}{t_2}\mathbf{v}_k$ is a linear combination of the others. □

Lemma 6.4.1 gives a way to enlarge independent sets to a basis; by contrast, Lemma 6.4.3 shows that spanning sets can be cut down to a basis.

Theorem 6.4.3

Let V be a finite dimensional vector space. Any spanning set for V can be cut down (by deleting vectors) to a basis of V.

Proof. Since V is finite dimensional, it has a finite spanning set S. Among all spanning sets contained in S, choose S_0 containing the smallest number of vectors. It suffices to show that S_0 is independent (then S_0 is a basis, proving the theorem). Suppose, on the contrary, that S_0 is not independent. Then, by Lemma 6.4.3, some vector $\mathbf{u} \in S_0$ is a linear combination of the set $S_1 = S_0 \setminus \{\mathbf{u}\}$ of vectors in S_0 other than $\mathbf{u}$. It follows that $\text{span}\ S_0 = \text{span}\ S_1$, that is, $V = \text{span}\ S_1$. But S_1 has fewer elements than S_0 so this contradicts the choice of S_0. Hence S_0 is independent after all. □

Note that, with Theorem 6.4.1, Theorem 6.4.3 completes the promised proof of Theorem 5.2.6 for the case $V = \mathbb{R}^n$.

Example 6.4.6

Find a basis of $\mathbf{P}_3$ in the spanning set $S = \{1,\ x+x^2,\ 2x-3x^2,\ 1+3x-2x^2,\ x^3\}$.

Solution. Since $\dim \mathbf{P}_3 = 4$, we must eliminate one polynomial from S. It cannot be x^3 because the span of the rest of S is contained in $\mathbf{P}_2$. But eliminating $1+3x-2x^2$ does leave a basis (verify). Note that $1+3x-2x^2$ is the sum of the first three polynomials in S.

Theorems 6.4.1 and 6.4.3 have other useful consequences.

Theorem 6.4.4

Let V be a vector space with $\dim V = n$, and suppose S is a set of exactly n vectors in V. Then S is independent if and only if S spans V.

Proof. Assume first that S is independent. By Theorem 6.4.1, S is contained in a basis B of V. Hence $|S| = n = |B|$ so, since $S \subseteq B$, it follows that $S = B$. In particular S spans V.

Conversely, assume that S spans V, so S contains a basis B by Theorem 6.4.3. Again $|S| = n = |B|$ so, since $S \supseteq B$, it follows that $S = B$. Hence S is independent. □

One of independence or spanning is often easier to establish than the other when showing that a set of vectors is a basis. For example if $V = \mathbb{R}^n$ it is easy to check whether a subset S of $\mathbb{R}^n$ is orthogonal (hence independent) but checking spanning can be tedious. Here are three more examples.

Example 6.4.7

Consider the set $S = \{p_0(x),\ p_1(x),\ \dots,\ p_n(x)\}$ of polynomials in $\mathbf{P}_n$. If $\deg p_k(x) = k$ for each k, show that S is a basis of $\mathbf{P}_n$.

Solution. The set S is independent—the degrees are distinct—see Example 6.3.4. Hence S is a basis of $\mathbf{P}_n$ by Theorem 6.4.4 because $\dim \mathbf{P}_n = n+1$.

Example 6.4.8

Let V denote the space of all symmetric 2×2 matrices. Find a basis of V consisting of invertible matrices.

Solution. We know that $\dim V = 3$ (Example 6.3.11), so what is needed is a set of three invertible, symmetric matrices that (using Theorem 6.4.4) is either independent or spans V. The set $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ is independent (verify) and so is a basis of the required type.

Example 6.4.9

Let A be any $n \times n$ matrix. Show that there exist n^2+1 scalars $a_0,\ a_1,\ a_2,\ \dots,\ a_{n^2}$ not all zero, such that

$$a_0 I + a_1 A + a_2 A^2 + \cdots + a_{n^2} A^{n^2} = 0$$

where I denotes the $n \times n$ identity matrix.

Solution. The space $\mathbf{M}_{nn}$ of all $n \times n$ matrices has dimension n^2 by Example 6.3.7. Hence the n^2+1 matrices $I,\ A,\ A^2,\ \dots,\ A^{n^2}$ cannot be independent by Theorem 6.4.4, so a nontrivial linear combination vanishes. This is the desired conclusion.

The result in Example 6.4.9 can be written as $f(A) = 0$ where $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n^2} x^{n^2}$. In other words, A satisfies a nonzero polynomial $f(x)$ of degree at most n^2. In fact we know that A satisfies a nonzero polynomial of degree n (this is the Cayley-Hamilton theorem—see Theorem 8.7.10), but the brevity of the solution in Example 6.4.6 is an indication of the power of these methods.

If U and W are subspaces of a vector space V, there are two related subspaces that are of interest, their **sum** $U+W$ and their **intersection** $U \cap W$, defined by

$$U+W = \{\mathbf{u}+\mathbf{w} \mid \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}$$
$$U \cap W = \{\mathbf{v} \in V \mid \mathbf{v} \in U \text{ and } \mathbf{v} \in W\}$$

It is routine to verify that these are indeed subspaces of V, that $U \cap W$ is contained in both U and W, and that $U+W$ contains both U and W. We conclude this section with a useful fact about the dimensions of these spaces. The proof is a good illustration of how the theorems in this section are used.

Theorem 6.4.5

Suppose that U and W are finite dimensional subspaces of a vector space V. Then $U+W$ is finite dimensional and

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W).$$

Proof. Since $U \cap W \subseteq U$, it has a finite basis, say $\{\mathbf{x}_1, \dots, \mathbf{x}_d\}$. Extend it to a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_d, \mathbf{u}_1, \dots, \mathbf{u}_m\}$ of U by Theorem 6.4.1. Similarly extend $\{\mathbf{x}_1, \dots, \mathbf{x}_d\}$ to a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_d, \mathbf{w}_1, \dots, \mathbf{w}_p\}$ of W. Then

$$U+W = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_d, \mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{w}_1, \dots, \mathbf{w}_p\}$$

as the reader can verify, so $U+W$ is finite dimensional. For the rest, it suffices to show that $\{\mathbf{x}_1, \dots, \mathbf{x}_d, \mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{w}_1, \dots, \mathbf{w}_p\}$ is independent (verify). Suppose that

$$r_1\mathbf{x}_1 + \cdots + r_d\mathbf{x}_d + s_1\mathbf{u}_1 + \cdots + s_m\mathbf{u}_m + t_1\mathbf{w}_1 + \cdots + t_p\mathbf{w}_p = \mathbf{0} \tag{6.1}$$

where the r_i, s_j, and t_k are scalars. Then

$$r_1\mathbf{x}_1 + \cdots + r_d\mathbf{x}_d + s_1\mathbf{u}_1 + \cdots + s_m\mathbf{u}_m = -(t_1\mathbf{w}_1 + \cdots + t_p\mathbf{w}_p)$$

is in U (left side) and also in W (right side), and so is in $U \cap W$. Hence $(t_1\mathbf{w}_1 + \cdots + t_p\mathbf{w}_p)$ is a linear combination of $\{\mathbf{x}_1, \dots, \mathbf{x}_d\}$, so $t_1 = \cdots = t_p = 0$, because $\{\mathbf{x}_1, \dots, \mathbf{x}_d, \mathbf{w}_1, \dots, \mathbf{w}_p\}$ is independent. Similarly, $s_1 = \cdots = s_m = 0$, so (6.1) becomes $r_1\mathbf{x}_1 + \cdots + r_d\mathbf{x}_d = \mathbf{0}$. It follows that $r_1 = \cdots = r_d = 0$, as required. □

Theorem 6.4.5 is particularly interesting if $U \cap W = \{\mathbf{0}\}$. Then there are *no* vectors $\mathbf{x}_i$ in the above proof, and the argument shows that if $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$ are bases of U and W respectively, then $\{\mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{w}_1, \dots, \mathbf{w}_p\}$ is a basis of $U+W$. In this case $U+W$ is said to be a **direct sum** (written $U \oplus W$); we return to this in Chapter 9.

Exercises for 6.4

Exercise 6.4.1 In each case, find a basis for V that includes the vector $\mathbf{v}$.

a. $V = \mathbb{R}^3$, $\mathbf{v} = (1, -1, 1)$

b. $V = \mathbb{R}^3$, $\mathbf{v} = (0, 1, 1)$

c. $V = \mathbf{M}_{22}$, $\mathbf{v} = \left[\begin{array}{rr} 1 & 1 \\ 1 & 1 \end{array}\right]$

d. $V = \mathbf{P}_2$, $\mathbf{v} = x^2 - x + 1$

Exercise 6.4.2 In each case, find a basis for V among the given vectors.

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

b. $V = \mathbf{P}_2$, $\{x^2 + 3, x + 2, x^2 - 2x - 1, x^2 + x\}$

Exercise 6.4.3 In each case, find a basis of V containing $\mathbf{v}$ and $\mathbf{w}$.

a. $V = \mathbb{R}^4$, $\mathbf{v} = (1, -1, 1, -1)$, $\mathbf{w} = (0, 1, 0, 1)$

b. $V = \mathbb{R}^4$, $\mathbf{v} = (0, 0, 1, 1)$, $\mathbf{w} = (1, 1, 1, 1)$

c. $V = \mathbf{M}_{22}$, $\mathbf{v} = \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right]$, $\mathbf{w} = \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \end{array}\right]$

d. $V = \mathbf{P}_3$, $\mathbf{v} = x^2 + 1$, $\mathbf{w} = x^2 + x$

Exercise 6.4.4

a. If z is not a real number, show that $\{z, z^2\}$ is a basis of the real vector space $\mathbb{C}$ of all complex numbers.

b. If z is neither real nor pure imaginary, show that $\{z, \overline{z}\}$ is a basis of $\mathbb{C}$.

Exercise 6.4.5 In each case use Theorem 6.4.4 to decide if S is a basis of V.

a. $V = \mathbf{M}_{22}$;
$S = \left\{ \left[\begin{array}{rr} 1 & 1 \\ 1 & 1 \end{array}\right], \left[\begin{array}{rr} 0 & 1 \\ 1 & 1 \end{array}\right], \left[\begin{array}{rr} 0 & 0 \\ 1 & 1 \end{array}\right], \left[\begin{array}{rr} 0 & 0 \\ 0 & 1 \end{array}\right] \right\}$

b. $V = \mathbf{P}_3$; $S = \{2x^2, 1 + x, 3, 1 + x + x^2 + x^3\}$

Exercise 6.4.6

a. Find a basis of $\mathbf{M}_{22}$ consisting of matrices with the property that $A^2 = A$.

b. Find a basis of $\mathbf{P}_3$ consisting of polynomials whose coefficients sum to 4. What if they sum to 0?

Exercise 6.4.7 If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis of V, determine which of the following are bases.

a. $\{\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{w}, \mathbf{v} + \mathbf{w}\}$

b. $\{2\mathbf{u} + \mathbf{v} + 3\mathbf{w}, 3\mathbf{u} + \mathbf{v} - \mathbf{w}, \mathbf{u} - 4\mathbf{w}\}$

c. $\{\mathbf{u}, \mathbf{u} + \mathbf{v} + \mathbf{w}\}$

d. $\{\mathbf{u}, \mathbf{u} + \mathbf{w}, \mathbf{u} - \mathbf{w}, \mathbf{v} + \mathbf{w}\}$

Exercise 6.4.8

a. Can two vectors span $\mathbb{R}^3$? Can they be linearly independent? Explain.

b. Can four vectors span $\mathbb{R}^3$? Can they be linearly independent? Explain.

Exercise 6.4.9 Show that any nonzero vector in a finite dimensional vector space is part of a basis.

Exercise 6.4.10 If A is a square matrix, show that $\det A = 0$ if and only if some row is a linear combination of the others.

Exercise 6.4.11 Let D, I, and X denote finite, nonempty sets of vectors in a vector space V. Assume that D is dependent and I is independent. In each case answer yes or no, and defend your answer.

a. If $X \supseteq D$, must X be dependent?

b. If $X \subseteq D$, must X be dependent?

c. If $X \supseteq I$, must X be independent?

d. If $X \subseteq I$, must X be independent?

Exercise 6.4.12 If U and W are subspaces of V and $\dim U = 2$, show that either $U \subseteq W$ or $\dim(U \cap W) \leq 1$.

Exercise 6.4.13 Let A be a nonzero 2×2 matrix and write $U = \{X \text{ in } \mathbf{M}_{22} \mid XA = AX\}$. Show that $\dim U \geq 2$. [*Hint*: I and A are in U.]

Exercise 6.4.14 If $U \subseteq \mathbb{R}^2$ is a subspace, show that $U = \{\mathbf{0}\}$, $U = \mathbb{R}^2$, or U is a line through the origin.

Exercise 6.4.15 Given $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k$, and $\mathbf{v}$, let $U = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ and $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}\}$. Show that either $\dim W = \dim U$ or $\dim W = 1 + \dim U$.

Exercise 6.4.16 Suppose U is a subspace of $\mathbf{P}_1$, $U \neq \{0\}$, and $U \neq \mathbf{P}_1$. Show that either $U = \mathbb{R}$ or $U = \mathbb{R}(a + x)$ for some a in $\mathbb{R}$.

Exercise 6.4.17 Let U be a subspace of V and assume $\dim V = 4$ and $\dim U = 2$. Does every basis of V result from adding (two) vectors to some basis of U? Defend your answer.

Exercise 6.4.18 Let U and W be subspaces of a vector space V.

a. If $\dim V = 3$, $\dim U = \dim W = 2$, and $U \neq W$, show that $\dim(U \cap W) = 1$.

b. Interpret (a.) geometrically if $V = \mathbb{R}^3$.

Exercise 6.4.19 Let $U \subseteq W$ be subspaces of V with $\dim U = k$ and $\dim W = m$, where $k < m$. If $k < l < m$, show that a subspace X exists where $U \subseteq X \subseteq W$ and $\dim X = l$.

Exercise 6.4.20 Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a *maximal* independent set in a vector space V. That is, no set of more than n vectors S is independent. Show that B is a basis of V.

Exercise 6.4.21 Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a *minimal* spanning set for a vector space V. That is, V cannot be spanned by fewer than n vectors. Show that B is a basis of V.

Exercise 6.4.22

a. Let $p(x)$ and $q(x)$ lie in $\mathbf{P}_1$ and suppose that $p(1) \neq 0$, $q(2) \neq 0$, and $p(2) = 0 = q(1)$. Show that $\{p(x), q(x)\}$ is a basis of $\mathbf{P}_1$. [*Hint*: If $rp(x) + sq(x) = 0$, evaluate at $x = 1$, $x = 2$.]

b. Let $B = \{p_0(x), p_1(x), \dots, p_n(x)\}$ be a set of polynomials in $\mathbf{P}_n$. Assume that there exist numbers $a_0, a_1, \dots, a_n$ such that $p_i(a_i) \neq 0$ for each i but $p_i(a_j) = 0$ if i is different from j. Show that B is a basis of $\mathbf{P}_n$.

Exercise 6.4.23 Let V be the set of all infinite sequences $(a_0, a_1, a_2, \dots)$ of real numbers. Define addition and scalar multiplication by

$$(a_0, a_1, \dots) + (b_0, b_1, \dots) = (a_0 + b_0, a_1 + b_1, \dots)$$

and

$$r(a_0, a_1, \dots) = (ra_0, ra_1, \dots)$$

a. Show that V is a vector space.

b. Show that V is not finite dimensional.

c. [For those with some calculus.] Show that the set of convergent sequences (that is, $\lim_{n\to\infty} a_n$ exists) is a subspace, also of infinite dimension.

Exercise 6.4.24 Let A be an $n \times n$ matrix of rank r. If $U = \{X \text{ in } \mathbf{M}_{nn} \mid AX = 0\}$, show that $\dim U = n(n - r)$. [*Hint*: Exercise 6.3.34.]

Exercise 6.4.25 Let U and W be subspaces of V.

a. Show that $U + W$ is a subspace of V containing both U and W.

b. Show that $\operatorname{span}\{\mathbf{u}, \mathbf{w}\} = \mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{w}$ for any vectors $\mathbf{u}$ and $\mathbf{w}$.

c. Show that

$$\begin{aligned} &\operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{w}_1, \dots, \mathbf{w}_n\} \\ &= \operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\} + \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_n\} \end{aligned}$$

for any vectors $\mathbf{u}_i$ in U and $\mathbf{w}_j$ in W.

Exercise 6.4.26 If A and B are $m \times n$ matrices, show that $\operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B$. [*Hint*: If U and V are the column spaces of A and B, respectively, show that the column space of $A + B$ is contained in $U + V$ and that $\dim(U + V) \leq \dim U + \dim V$. (See Theorem 6.4.5.)]

6.5 An Application to Polynomials

The vector space of all polynomials of degree at most n is denoted $\mathbf{P}_n$, and it was established in Section 6.3 that $\mathbf{P}_n$ has dimension $n+1$; in fact, $\{1, x, x^2, \dots, x^n\}$ is a basis. More generally, *any* $n+1$ polynomials of distinct degrees form a basis, by Theorem 6.4.4 (they are independent by Example 6.3.4). This proves

Theorem 6.5.1

Let $p_0(x), p_1(x), p_2(x), \dots, p_n(x)$ be polynomials in $\mathbf{P}_n$ of degrees $0, 1, 2, \dots, n$, respectively. Then $\{p_0(x), \dots, p_n(x)\}$ is a basis of $\mathbf{P}_n$.

An immediate consequence is that $\{1, (x-a), (x-a)^2, \dots, (x-a)^n\}$ is a basis of $\mathbf{P}_n$ for any number a. Hence we have the following:

Corollary 6.5.1

If a is any number, every polynomial $f(x)$ of degree at most n has an expansion in powers of $(x-a)$:

$$f(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + \cdots + a_n(x-a)^n \tag{6.2}$$

If $f(x)$ is evaluated at $x = a$, then equation (6.2) becomes

$$f(x) = a_0 + a_1(a-a) + \cdots + a_n(a-a)^n = a_0$$

Hence $a_0 = f(a)$, and equation (6.2) can be written $f(x) = f(a) + (x-a)g(x)$, where $g(x)$ is a polynomial of degree $n-1$ (this assumes that $n \geq 1$). If it happens that $f(a) = 0$, then it is clear that $f(x)$ has the form $f(x) = (x-a)g(x)$. Conversely, every such polynomial certainly satisfies $f(a) = 0$, and we obtain:

Corollary 6.5.2

Let $f(x)$ be a polynomial of degree $n \geq 1$ and let a be any number. Then:

Remainder Theorem

1. *$f(x) = f(a) + (x-a)g(x)$ for some polynomial $g(x)$ of degree $n-1$.*

Factor Theorem

2. *$f(a) = 0$ if and only if $f(x) = (x-a)g(x)$ for some polynomial $g(x)$.*

The polynomial $g(x)$ can be computed easily by using "long division" to divide $f(x)$ by $(x-a)$—see Appendix D.

All the coefficients in the expansion (6.2) of $f(x)$ in powers of $(x-a)$ can be determined in terms of the derivatives of $f(x)$.[6] These will be familiar to students of calculus. Let $f^{(n)}(x)$ denote the nth derivative

[6]The discussion of Taylor's theorem can be omitted with no loss of continuity.

of the polynomial $f(x)$, and write $f^{(0)}(x) = f(x)$. Then, if

$$f(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + \cdots + a_n(x-a)^n$$

it is clear that $a_0 = f(a) = f^{(0)}(a)$. Differentiation gives

$$f^{(1)}(x) = a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + \cdots + na_n(x-a)^{n-1}$$

and substituting $x = a$ yields $a_1 = f^{(1)}(a)$. This continues to give $a_2 = \frac{f^{(2)}(a)}{2!}$, $a_3 = \frac{f^{(3)}(a)}{3!}$, $\dots$, $a_k = \frac{f^{(k)}(a)}{k!}$, where $k!$ is defined as $k! = k(k-1)\cdots 2 \cdot 1$. Hence we obtain the following:

Corollary 6.5.3: Taylor's Theorem

If $f(x)$ is a polynomial of degree n, then

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Example 6.5.1

Expand $f(x) = 5x^3 + 10x + 2$ as a polynomial in powers of $x-1$.

Solution. The derivatives are $f^{(1)}(x) = 15x^2 + 10$, $f^{(2)}(x) = 30x$, and $f^{(3)}(x) = 30$. Hence the Taylor expansion is

$$\begin{aligned} f(x) &= f(1) + \frac{f^{(1)}(1)}{1!}(x-1) + \frac{f^{(2)}(1)}{2!}(x-1)^2 + \frac{f^{(3)}(1)}{3!}(x-1)^3 \\ &= 17 + 25(x-1) + 15(x-1)^2 + 5(x-1)^3 \end{aligned}$$

Taylor's theorem is useful in that it provides a formula for the coefficients in the expansion. It is dealt with in calculus texts and will not be pursued here.

Theorem 6.5.1 produces bases of $\mathbf{P}_n$ consisting of polynomials of distinct degrees. A different criterion is involved in the next theorem.

Theorem 6.5.2

Let $f_0(x), f_1(x), \dots, f_n(x)$ be nonzero polynomials in $\mathbf{P}_n$. Assume that numbers $a_0, a_1, \dots, a_n$ exist such that

$$\begin{aligned} f_i(a_i) &\neq 0 \quad \text{for each } i \\ f_i(a_j) &= 0 \quad \text{if } i \neq j \end{aligned}$$

Then

1. *$\{f_0(x), \dots, f_n(x)\}$ is a basis of $\mathbf{P}_n$.*
2. *If $f(x)$ is any polynomial in $\mathbf{P}_n$, its expansion as a linear combination of these basis vectors is*

$$f(x) = \frac{f(a_0)}{f_0(a_0)}f_0(x) + \frac{f(a_1)}{f_1(a_1)}f_1(x) + \cdots + \frac{f(a_n)}{f_n(a_n)}f_n(x)$$

Proof.

1. It suffices (by Theorem 6.4.4) to show that $\{f_0(x), \dots, f_n(x)\}$ is linearly independent (because $\dim \mathbf{P}_n = n+1$). Suppose that

$$r_0 f_0(x) + r_1 f_1(x) + \cdots + r_n f_n(x) = 0, \ r_i \in \mathbb{R}$$

Because $f_i(a_0) = 0$ for all $i > 0$, taking $x = a_0$ gives $r_0 f_0(a_0) = 0$. But then $r_0 = 0$ because $f_0(a_0) \neq 0$. The proof that $r_i = 0$ for $i > 0$ is analogous.

2. By (1), $f(x) = r_0 f_0(x) + \cdots + r_n f_n(x)$ for *some* numbers r_i. Once again, evaluating at a_0 gives $f(a_0) = r_0 f_0(a_0)$, so $r_0 = f(a_0)/f_0(a_0)$. Similarly, $r_i = f(a_i)/f_i(a_i)$ for each i. □

Example 6.5.2

Show that $\{x^2 - x, \ x^2 - 2x, \ x^2 - 3x + 2\}$ is a basis of $\mathbf{P}_2$.

Solution. Write $f_0(x) = x^2 - x = x(x-1)$, $f_1(x) = x^2 - 2x = x(x-2)$, and $f_2(x) = x^2 - 3x + 2 = (x-1)(x-2)$. Then the conditions of Theorem 6.5.2 are satisfied with $a_0 = 2$, $a_1 = 1$, and $a_2 = 0$.

We investigate one natural choice of the polynomials $f_i(x)$ in Theorem 6.5.2. To illustrate, let a_0, a_1, and a_2 be distinct numbers and write

$$f_0(x) = \frac{(x-a_1)(x-a_2)}{(a_0-a_1)(a_0-a_2)} \qquad f_1(x) = \frac{(x-a_0)(x-a_2)}{(a_1-a_0)(a_1-a_2)} \qquad f_2(x) = \frac{(x-a_0)(x-a_1)}{(a_2-a_0)(a_2-a_1)}$$

Then $f_0(a_0) = f_1(a_1) = f_2(a_2) = 1$, and $f_i(a_j) = 0$ for $i \neq j$. Hence Theorem 6.5.2 applies, and because $f_i(a_i) = 1$ for each i, the formula for expanding any polynomial is simplified.

In fact, this can be generalized with no extra effort. If $a_0, a_1, \dots, a_n$ are distinct numbers, define the **Lagrange polynomials** $\delta_0(x), \delta_1(x), \dots, \delta_n(x)$ relative to these numbers as follows:

$$\delta_k(x) = \frac{\prod_{i \neq k}(x-a_i)}{\prod_{i \neq k}(a_k-a_i)} \qquad k = 0, 1, 2, \dots, n$$

Here the numerator is the product of all the terms $(x-a_0), (x-a_1), \dots, (x-a_n)$ with $(x-a_k)$ omitted, and a similar remark applies to the denominator. If $n = 2$, these are just the polynomials in the preceding paragraph. For another example, if $n = 3$, the polynomial $\delta_1(x)$ takes the form

$$\delta_1(x) = \frac{(x-a_0)(x-a_2)(x-a_3)}{(a_1-a_0)(a_1-a_2)(a_1-a_3)}$$

In the general case, it is clear that $\delta_i(a_i) = 1$ for each i and that $\delta_i(a_j) = 0$ if $i \neq j$. Hence Theorem 6.5.2 specializes as Theorem 6.5.3.

Theorem 6.5.3: Lagrange Interpolation Expansion

Let $a_0,\ a_1,\ \ldots,\ a_n$ be distinct numbers. The corresponding set

$$\{\delta_0(x),\ \delta_1(x),\ \ldots,\ \delta_n(x)\}$$

of Lagrange polynomials is a basis of $\mathbf{P}_n$, and any polynomial $f(x)$ in $\mathbf{P}_n$ has the following unique expansion as a linear combination of these polynomials.

$$f(x) = f(a_0)\delta_0(x) + f(a_1)\delta_1(x) + \cdots + f(a_n)\delta_n(x)$$

Example 6.5.3

Find the Lagrange interpolation expansion for $f(x) = x^2 - 2x + 1$ relative to $a_0 = -1$, $a_1 = 0$, and $a_2 = 1$.

Solution. The Lagrange polynomials are

$$\begin{aligned}
\delta_0 &= \tfrac{(x-0)(x-1)}{(-1-0)(-1-1)} = \tfrac{1}{2}(x^2 - x) \\
\delta_1 &= \tfrac{(x+1)(x-1)}{(0+1)(0-1)} = -(x^2 - 1) \\
\delta_2 &= \tfrac{(x+1)(x-0)}{(1+1)(1-0)} = \tfrac{1}{2}(x^2 + x)
\end{aligned}$$

Because $f(-1) = 4$, $f(0) = 1$, and $f(1) = 0$, the expansion is

$$f(x) = 2(x^2 - x) - (x^2 - 1)$$

The Lagrange interpolation expansion gives an easy proof of the following important fact.

Theorem 6.5.4

Let $f(x)$ be a polynomial in $\mathbf{P}_n$, and let $a_0,\ a_1,\ \ldots,\ a_n$ denote distinct numbers. If $f(a_i) = 0$ for all i, then $f(x)$ is the zero polynomial (that is, all coefficients are zero).

Proof. All the coefficients in the Lagrange expansion of $f(x)$ are zero. □

Exercises for 6.5

Exercise 6.5.1 If polynomials $f(x)$ and $g(x)$ satisfy $f(a) = g(a)$, show that $f(x) - g(x) = (x-a)h(x)$ for some polynomial $h(x)$.

Exercises 6.5.2, 6.5.3, 6.5.4, and 6.5.5 require polynomial differentiation.

Exercise 6.5.2 Expand each of the following as a polynomial in powers of $x-1$.

a. $f(x) = x^3 - 2x^2 + x - 1$

b. $f(x) = x^3 + x + 1$

c. $f(x) = x^4$

d. $f(x) = x^3 - 3x^2 + 3x$

Exercise 6.5.3 Prove Taylor's theorem for polynomials.

Exercise 6.5.4 Use Taylor's theorem to derive the **binomial theorem**:

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$$

Here the **binomial coefficients** $\binom{n}{r}$ are defined by

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

where $n! = n(n-1)\cdots 2 \cdot 1$ if $n \geq 1$ and $0! = 1$.

Exercise 6.5.5 Let $f(x)$ be a polynomial of degree n. Show that, given any polynomial $g(x)$ in $\mathbf{P}_n$, there exist numbers $b_0, b_1, \dots, b_n$ such that

$$g(x) = b_0 f(x) + b_1 f^{(1)}(x) + \cdots + b_n f^{(n)}(x)$$

where $f^{(k)}(x)$ denotes the kth derivative of $f(x)$.

Exercise 6.5.6 Use Theorem 6.5.2 to show that the following are bases of $\mathbf{P}_2$.

a. $\{x^2 - 2x, x^2 + 2x, x^2 - 4\}$

b. $\{x^2 - 3x + 2, x^2 - 4x + 3, x^2 - 5x + 6\}$

Exercise 6.5.7 Find the Lagrange interpolation expansion of $f(x)$ relative to $a_0 = 1$, $a_1 = 2$, and $a_2 = 3$ if:

a. $f(x) = x^2 + 1$

b. $f(x) = x^2 + x + 1$

Exercise 6.5.8 Let $a_0, a_1, \dots, a_n$ be distinct numbers. If $f(x)$ and $g(x)$ in $\mathbf{P}_n$ satisfy $f(a_i) = g(a_i)$ for all i, show that $f(x) = g(x)$. [*Hint*: See Theorem 6.5.4.]

Exercise 6.5.9 Let $a_0, a_1, \dots, a_n$ be distinct numbers. If $f(x) \in \mathbf{P}_{n+1}$ satisfies $f(a_i) = 0$ for each $i = 0, 1, \dots, n$, show that $f(x) = r(x-a_0)(x-a_1)\cdots(x-a_n)$ for some r in $\mathbb{R}$. [*Hint*: r is the coefficient of x^{n+1} in $f(x)$. Consider $f(x) - r(x-a_0)\cdots(x-a_n)$ and use Theorem 6.5.4.]

Exercise 6.5.10 Let a and b denote distinct numbers.

a. Show that $\{(x-a), (x-b)\}$ is a basis of $\mathbf{P}_1$.

b. Show that $\{(x-a)^2, (x-a)(x-b), (x-b)^2\}$ is a basis of $\mathbf{P}_2$.

c. Show that $\{(x-a)^n, (x-a)^{n-1}(x-b), \dots, (x-a)(x-b)^{n-1}, (x-b)^n\}$ is a basis of $\mathbf{P}_n$. [*Hint*: If a linear combination vanishes, evaluate at $x = a$ and $x = b$. Then reduce to the case $n-2$ by using the fact that if $p(x)q(x) = 0$ in $\mathbf{P}$, then either $p(x) = 0$ or $q(x) = 0$.]

Exercise 6.5.11 Let a and b be two distinct numbers. Assume that $n \geq 2$ and let

$$U_n = \{f(x) \text{ in } \mathbf{P}_n \mid f(a) = 0 = f(b)\}.$$

a. Show that

$$U_n = \{(x-a)(x-b)p(x) \mid p(x) \text{ in } \mathbf{P}_{n-2}\}$$

b. Show that $\dim U_n = n-1$.

[*Hint*: If $p(x)q(x) = 0$ in $\mathbf{P}$, then either $p(x) = 0$, or $q(x) = 0$.]

c. Show $\{(x-a)^{n-1}(x-b), (x-a)^{n-2}(x-b)^2, \dots, (x-a)^2(x-b)^{n-2}, (x-a)(x-b)^{n-1}\}$ is a basis of U_n. [*Hint*: Exercise 6.5.10.]

6.6 An Application to Differential Equations

Call a function $f : \mathbb{R} \to \mathbb{R}$ **differentiable** if it can be differentiated as many times as we want. If f is a differentiable function, the nth derivative $f^{(n)}$ of f is the result of differentiating n times. Thus $f^{(0)} = f$, $f^{(1)} = f'$, $f^{(2)} = f^{(1)\prime}$, $\dots$ and, in general, $f^{(n+1)} = f^{(n)\prime}$ for each $n \geq 0$. For small values of n these are often written as $f, f', f'', f''', \dots$.

If a, b, and c are numbers, the differential equations

$$f'' + af' + bf = 0 \quad \text{or} \quad f''' + af'' + bf' + cf = 0$$

are said to be of **second-order** and **third-order**, respectively. In general, an equation

$$f^{(n)} + a_{n-1}f^{(n-1)} + a_{n-2}f^{(n-2)} + \cdots + a_2f^{(2)} + a_1f^{(1)} + a_0f^{(0)} = 0, \quad a_i \text{ in } \mathbb{R} \tag{6.3}$$

is called a **differential equation of order** n. In this section we investigate the set of solutions to (6.3) and, if n is 1 or 2, find explicit solutions. Of course an acquaintance with calculus is required.

Let f and g be solutions to (6.3). Then $f+g$ is also a solution because $(f+g)^{(k)} = f^{(k)} + g^{(k)}$ for all k, and af is a solution for any a in $\mathbb{R}$ because $(af)^{(k)} = af^{(k)}$. It follows that the set of solutions to (6.3) is a vector space, and we ask for the dimension of this space.

We have already dealt with the simplest case (see Theorem 3.5.1):

Theorem 6.6.1

The set of solutions of the first-order differential equation $f' + af = 0$ is a one-dimensional vector space and $\{e^{-ax}\}$ is a basis.

There is a far-reaching generalization of Theorem 6.6.1 that will be proved in Theorem 7.4.1.

Theorem 6.6.2

The set of solutions to the nth order equation (6.3) has dimension n.

Remark

Every differential equation of order n can be converted into a system of n linear first-order equations (see Exercises 3.5.6 and 3.5.7). In the case that the matrix of this system is diagonalizable, this approach provides a proof of Theorem 6.6.2. But if the matrix is not diagonalizable, Theorem 7.4.1 is required.

Theorem 6.6.1 suggests that we look for solutions to (6.3) of the form $e^{\lambda x}$ for some number λ. This is a good idea. If we write $f(x) = e^{\lambda x}$, it is easy to verify that $f^{(k)}(x) = \lambda^k e^{\lambda x}$ for each $k \geq 0$, so substituting f in (6.3) gives

$$(\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_2\lambda^2 + a_1\lambda^1 + a_0)e^{\lambda x} = 0$$

Since $e^{\lambda x} \neq 0$ for all x, this shows that $e^{\lambda x}$ is a solution of (6.3) if and only if λ is a root of the **characteristic polynomial** $c(x)$, defined to be

$$c(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0$$

This proves Theorem 6.6.3.

Theorem 6.6.3

If λ is real, the function $e^{\lambda x}$ is a solution of (6.3) if and only if λ is a root of the characteristic polynomial $c(x)$.

Example 6.6.1

Find a basis of the space U of solutions of $f''' - 2f'' - f' - 2f = 0$.

Solution. The characteristic polynomial is $x^3 - 2x^2 - x - 1 = (x-1)(x+1)(x-2)$, with roots $\lambda_1 = 1$, $\lambda_2 = -1$, and $\lambda_3 = 2$. Hence e^x, e^{-x}, and e^{2x} are all in U. Moreover they are independent (by Lemma 6.6.1 below) so, since $\dim(U) = 3$ by Theorem 6.6.2, $\{e^x, e^{-x}, e^{2x}\}$ is a basis of U.

Lemma 6.6.1

If $\lambda_1, \lambda_2, \dots, \lambda_k$ are distinct, then $\{e^{\lambda_1 x}, e^{\lambda_2 x}, \dots, e^{\lambda_k x}\}$ is linearly independent.

Proof. If $r_1 e^{\lambda_1 x} + r_2 e^{\lambda_2 x} + \cdots + r_k e^{\lambda_k x} = 0$ for all x, then $r_1 + r_2 e^{(\lambda_2 - \lambda_1)x} + \cdots + r_k e^{(\lambda_k - \lambda_1)x} = 0$; that is, $r_2 e^{(\lambda_2 - \lambda_1)x} + \cdots + r_k e^{(\lambda_k - \lambda_1)x}$ is a constant. Since the λ_i are distinct, this forces $r_2 = \cdots = r_k = 0$, whence $r_1 = 0$ also. This is what we wanted. □

Theorem 6.6.4

Let U denote the space of solutions to the second-order equation

$$f'' + af' + bf = 0$$

where a and b are real constants. Assume that the characteristic polynomial $x^2 + ax + b$ has two real roots λ and μ. Then

1. *If $\lambda \neq \mu$, then $\{e^{\lambda x}, e^{\mu x}\}$ is a basis of U.*
2. *If $\lambda = \mu$, then $\{e^{\lambda x}, xe^{\lambda x}\}$ is a basis of U.*

Proof. Since $\dim(U) = 2$ by Theorem 6.6.2, (1) follows by Lemma 6.6.1, and (2) follows because the set $\{e^{\lambda x}, xe^{\lambda x}\}$ is independent (Exercise 6.6.3). □

Example 6.6.2

Find the solution of $f'' + 4f' + 4f = 0$ that satisfies the **boundary conditions** $f(0) = 1$, $f(1) = -1$.

Solution. The characteristic polynomial is $x^2+4x+4=(x+2)^2$, so -2 is a double root. Hence $\{e^{-2x},\ xe^{-2x}\}$ is a basis for the space of solutions, and the general solution takes the form $f(x)=ce^{-2x}+dxe^{-2x}$. Applying the boundary conditions gives $1=f(0)=c$ and $-1=f(1)=(c+d)e^{-2}$. Hence $c=1$ and $d=-(1+e^2)$, so the required solution is

$$f(x)=e^{-2x}-(1+e^2)xe^{-2x}$$

One other question remains: What happens if the roots of the characteristic polynomial are not real? To answer this, we must first state precisely what $e^{\lambda x}$ means when λ is not real. If q is a real number, define

$$e^{iq}=\cos q+i\sin q$$

where $i^2=-1$. Then the relationship $e^{iq}e^{iq_1}=e^{i(q+q_1)}$ holds for all real q and q_1, as is easily verified. If $\lambda=p+iq$, where p and q are real numbers, we define

$$e^{\lambda}=e^{p}e^{iq}=e^{p}(\cos q+i\sin q)$$

Then it is a routine exercise to show that

1. $e^{\lambda}e^{\mu}=e^{\lambda+\mu}$
2. $e^{\lambda}=1$ if and only if $\lambda=0$
3. $(e^{\lambda x})'=\lambda e^{\lambda x}$

These easily imply that $f(x)=e^{\lambda x}$ is a solution to $f''+af'+bf=0$ if λ is a (possibly complex) root of the characteristic polynomial x^2+ax+b. Now write $\lambda=p+iq$ so that

$$f(x)=e^{\lambda x}=e^{px}\cos(qx)+ie^{px}\sin(qx)$$

For convenience, denote the real and imaginary parts of $f(x)$ as $u(x)=e^{px}\cos(qx)$ and $v(x)=e^{px}\sin(qx)$. Then the fact that $f(x)$ satisfies the differential equation gives

$$0=f''+af'+bf=(u''+au'+bu)+i(v''+av'+bv)$$

Equating real and imaginary parts shows that $u(x)$ and $v(x)$ are both solutions to the differential equation. This proves part of Theorem 6.6.5.

Theorem 6.6.5

Let U denote the space of solutions of the second-order differential equation

$$f''+af'+bf=0$$

where a and b are real. Suppose λ is a nonreal root of the characteristic polynomial x^2+ax+b. If $\lambda=p+iq$, where p and q are real, then

$$\{e^{px}\cos(qx),\ e^{px}\sin(qx)\}$$

is a basis of U.

Proof. The foregoing discussion shows that these functions lie in U. Because $\dim U = 2$ by Theorem 6.6.2, it suffices to show that they are linearly independent. But if

$$re^{px}\cos(qx) + se^{px}\sin(qx) = 0$$

for all x, then $r\cos(qx) + s\sin(qx) = 0$ for all x (because $e^{px} \neq 0$). Taking $x = 0$ gives $r = 0$, and taking $x = \frac{\pi}{2q}$ gives $s = 0$ ($q \neq 0$ because λ is not real). This is what we wanted. □

Example 6.6.3

Find the solution $f(x)$ to $f'' - 2f' + 2f = 0$ that satisfies $f(0) = 2$ and $f(\frac{\pi}{2}) = 0$.

Solution. The characteristic polynomial $x^2 - 2x + 2$ has roots $1 + i$ and $1 - i$. Taking $\lambda = 1 + i$ (quite arbitrarily) gives $p = q = 1$ in the notation of Theorem 6.6.5, so $\{e^x \cos x,\ e^x \sin x\}$ is a basis for the space of solutions. The general solution is thus $f(x) = e^x(r\cos x + s\sin x)$. The boundary conditions yield $2 = f(0) = r$ and $0 = f(\frac{\pi}{2}) = e^{\pi/2}s$. Thus $r = 2$ and $s = 0$, and the required solution is $f(x) = 2e^x \cos x$.

The following theorem is an important special case of Theorem 6.6.5.

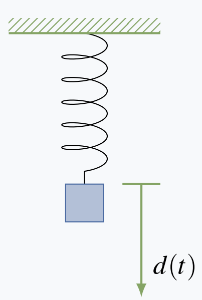
Theorem 6.6.6

If $q \neq 0$ is a real number, the space of solutions to the differential equation $f'' + q^2 f = 0$ has basis $\{\cos(qx),\ \sin(qx)\}$.

Proof. The characteristic polynomial $x^2 + q^2$ has roots qi and $-qi$, so Theorem 6.6.5 applies with $p = 0$. □

In many situations, the displacement $s(t)$ of some object at time t turns out to have an oscillating form $s(t) = c\sin(at) + d\cos(at)$. These are called **simple harmonic motions**. An example follows.

Example 6.6.4



A weight is attached to an extension spring (see diagram). If it is pulled from the equilibrium position and released, it is observed to oscillate up and down. Let $d(t)$ denote the distance of the weight below the equilibrium position t seconds later. It is known (**Hooke's law**) that the acceleration $d''(t)$ of the weight is proportional to the displacement $d(t)$ and in the opposite direction. That is,

$$d''(t) = -kd(t)$$

where $k > 0$ is called the **spring constant**. Find $d(t)$ if the maximum extension is 10 cm below the equilibrium position and find the **period** of the oscillation (time taken for the weight to make a full oscillation).

Solution. It follows from Theorem 6.6.6 (with $q^2 = k$) that

$$d(t) = r\sin(\sqrt{k}\,t) + s\cos(\sqrt{k}\,t)$$

where r and s are constants. The condition $d(0) = 0$ gives $s = 0$, so $d(t) = r\sin(\sqrt{k}\,t)$. Now the maximum value of the function $\sin x$ is 1 (when $x = \frac{\pi}{2}$), so $r = 10$ (when $t = \frac{\pi}{2\sqrt{k}}$). Hence

$$d(t) = 10\sin(\sqrt{k}\,t)$$

Finally, the weight goes through a full oscillation as $\sqrt{k}\,t$ increases from 0 to 2π. The time taken is $t = \frac{2\pi}{\sqrt{k}}$, the period of the oscillation.

Exercises for 6.6

Exercise 6.6.1 Find a solution f to each of the following differential equations satisfying the given boundary conditions.

a. $f' - 3f = 0$; $f(1) = 2$

b. $f' + f = 0$; $f(1) = 1$

c. $f'' + 2f' - 15f = 0$; $f(1) = f(0) = 0$

d. $f'' + f' - 6f = 0$; $f(0) = 0$, $f(1) = 1$

e. $f'' - 2f' + f = 0$; $f(1) = f(0) = 1$

f. $f'' - 4f' + 4f = 0$; $f(0) = 2$, $f(-1) = 0$

g. $f'' - 3af' + 2a^2 f = 0$; $a \neq 0$; $f(0) = 0$, $f(1) = 1 - e^a$

h. $f'' - a^2 f = 0$, $a \neq 0$; $f(0) = 1$, $f(1) = 0$

i. $f'' - 2f' + 5f = 0$; $f(0) = 1$, $f(\frac{\pi}{4}) = 0$

j. $f'' + 4f' + 5f = 0$; $f(0) = 0$, $f(\frac{\pi}{2}) = 1$

Exercise 6.6.2 If the characteristic polynomial of $f'' + af' + bf = 0$ has real roots, show that $f = 0$ is the only solution satisfying $f(0) = 0 = f(1)$.

Exercise 6.6.3 Complete the proof of Theorem 6.6.2. [*Hint*: If λ is a double root of $x^2 + ax + b$, show that $a = -2\lambda$ and $b = \lambda^2$. Hence $xe^{\lambda x}$ is a solution.]

Exercise 6.6.4

a. Given the equation $f' + af = b$, $(a \neq 0)$, make the substitution $f(x) = g(x) + b/a$ and obtain a differential equation for g. Then derive the general solution for $f' + af = b$.

b. Find the general solution to $f' + f = 2$.

Exercise 6.6.5 Consider the differential equation $f' + af' + bf = g$, where g is some fixed function. Assume that f_0 is one solution of this equation.

a. Show that the general solution is $cf_1 + df_2 + f_0$, where c and d are constants and $\{f_1,\ f_2\}$ is any basis for the solutions to $f'' + af' + bf = 0$.

b. Find a solution to $f'' + f' - 6f = 2x^3 - x^2 - 2x$. [*Hint*: Try $f(x) = \frac{-1}{3}x^3$.]

Exercise 6.6.6 A radioactive element decays at a rate proportional to the amount present. Suppose an initial mass of 10 grams decays to 8 grams in 3 hours.

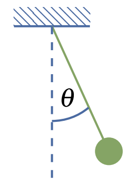
a. Find the mass t hours later.

b. Find the *half-life* of the element—the time it takes to decay to half its mass.

Exercise 6.6.7 The population $N(t)$ of a region at time t increases at a rate proportional to the population. If the population doubles in 5 years and is 3 million initially, find $N(t)$.

Exercise 6.6.8 Consider a spring, as in Example 6.6.4. If the period of the oscillation is 30 seconds, find the spring constant k.

Exercise 6.6.9 As a pendulum swings (see the diagram), let t measure the time since it was vertical. The angle $\theta = \theta(t)$ from the vertical can be shown to satisfy the equation $\theta'' + k\theta = 0$, provided that θ is small. If the maximal angle is $\theta = 0.05$ radians, find $\theta(t)$ in terms of k. If the period is 0.5 seconds, find k. [Assume that $\theta = 0$ when $t = 0$.]



Supplementary Exercises for Chapter 6

Exercise 6.1 (Requires calculus) Let V denote the space of all functions $f : \mathbb{R} \to \mathbb{R}$ for which the derivatives f' and f'' exist. Show that f_1, f_2, and f_3 in V are linearly independent provided that their **wronskian** $w(x)$ is nonzero for some x, where

$$w(x) = \det \begin{bmatrix} f_1(x) & f_2(x) & f_3(x) \\ f_1'(x) & f_2'(x) & f_3'(x) \\ f_1''(x) & f_2''(x) & f_3''(x) \end{bmatrix}$$

Exercise 6.2 Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis of $\mathbb{R}^n$ (written as columns), and let A be an $n \times n$ matrix.

a. If A is invertible, show that $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$.

b. If $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$, show that A is invertible.

Exercise 6.3 If A is an $m \times n$ matrix, show that A has rank m if and only if col A contains every column of I_m.

Exercise 6.4 Show that null A = null (A^TA) for any real matrix A.

Exercise 6.5 Let A be an $m \times n$ matrix of rank r. Show that $\dim(\text{null } A) = n - r$ (Theorem 5.4.3) as follows. Choose a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ of null A and extend it to a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{z}_1, \dots, \mathbf{z}_m\}$ of $\mathbb{R}^n$. Show that $\{A\mathbf{z}_1, \dots, A\mathbf{z}_m\}$ is a basis of col A.

Math 2331 – Linear Algebra

4.1 Vector Spaces & Subspaces

Jiwen He

Department of Mathematics, University of Houston

jiwenhe@math.uh.edu
math.uh.edu/∼jiwenhe/math2331

4.1 Vector Spaces & Subspaces

- Vector Spaces: Definition
- Vector Spaces: Examples
  - 2×2 matrices
  - Polynomials
- Subspaces: Definition
- Subspaces: Examples
- Determining Subspaces

Vector Spaces

Many concepts concerning vectors in $\mathbf{R}^n$ can be extended to other mathematical systems.

We can think of a *vector space* in general, as a collection of objects that behave as vectors do in $\mathbf{R}^n$. The objects of such a set are called *vectors*.

Vector Space

A **vector space** is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms below. The axioms must hold for all $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in V and for all scalars c and d.

1. $\mathbf{u} + \mathbf{v}$ is in V.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

Vector Spaces (cont.)

Vector Space (cont.)

3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
4. There is a vector (called the zero vector) $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is vector $-\mathbf{u}$ in V satisfying $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.
6. $c\mathbf{u}$ is in V.
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.
9. $(cd)\mathbf{u}=c(d\mathbf{u})$.
10. $1\mathbf{u}=\mathbf{u}$.

Vector Spaces: Examples

Example

Let $M_{2\times 2} = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a,\ b,\ c,\ d \text{ are real} \right\}$

In this context, note that the $\mathbf{0}$ vector is $\begin{bmatrix} \quad & \quad \end{bmatrix}$.

Vector Spaces: Polynomials

Example

Let $n \geq 0$ be an integer and let

$$\mathbf{P}_n = \text{the set of all polynomials of degree at most } n \geq 0.$$

Members of $\mathbf{P}_n$ have the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$$

where $a_0, a_1, \ldots, a_n$ are real numbers and t is a real variable. The set $\mathbf{P}_n$ is a vector space.

We will just verify 3 out of the 10 axioms here.

Let $\mathbf{p}(t) = a_0 + a_1 t + \cdots + a_n t^n$ and $\mathbf{q}(t) = b_0 + b_1 t + \cdots + b_n t^n$. Let c be a scalar.

Vector Spaces: Polynomials (cont.)

Axiom 1:

The polynomial $\mathbf{p}+\mathbf{q}$ is defined as follows:
$(\mathbf{p}+\mathbf{q})(t)=\mathbf{p}(t)+\mathbf{q}(t)$. Therefore,

$$(\mathbf{p}+\mathbf{q})(t)=\mathbf{p}(t)+\mathbf{q}(t)$$

$$=(\_\_\_\_\_\_\_\_\_)+(\_\_\_\_\_\_\_\_\_)\,t+\cdots+(\_\_\_\_\_\_\_\_\_)\,t^n$$

which is also a ____________________ of degree at most ________. So

$\mathbf{p}+\mathbf{q}$ is in $\mathbf{P}_n$.

Vector Spaces: Polynomials (cont.)

Axiom 4:

$$\mathbf{0} = 0 + 0t + \cdots + 0t^n$$

(zero vector in $\mathbf{P}_n$)

$$(\mathbf{p} + \mathbf{0})(t) = \mathbf{p}(t) + \mathbf{0} = (a_0 + 0) + (a_1 + 0)t + \cdots + (a_n + 0)t^n$$

$$= a_0 + a_1 t + \cdots + a_n t^n = \mathbf{p}(t)$$

and so $\mathbf{p} + \mathbf{0} = \mathbf{p}$

Vector Spaces: Polynomials (cont.)

Axiom 6:

$$(c\mathbf{p})(t) = c\mathbf{p}(t) = (\_\_\_\_\_\_\_\_) + (\_\_\_\_\_\_\_\_)\, t + \cdots + (\_\_\_\_\_\_\_\_)\, t^n$$

which is in $\mathbf{P}_n$.

The other 7 axioms also hold, so $\mathbf{P}_n$ is a vector space.

Subspaces

Vector spaces may be formed from subsets of other vectors spaces. These are called *subspaces*.

Subspaces

A **subspace** of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.

b. For each $\mathbf{u}$ and $\mathbf{v}$ are in H, $\mathbf{u}+\mathbf{v}$ is in H. (In this case we say H is closed under vector addition.)

c. For each $\mathbf{u}$ in H and each scalar c, $c\mathbf{u}$ is in H. (In this case we say H is closed under scalar multiplication.)

If the subset H satisfies these three properties, then H itself is a vector space.

Subspaces: Example

Example

Let $H = \left\{ \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} : a \text{ and } b \text{ are real} \right\}$. Show that H is a subspace of $\mathbf{R}^3$.

Solution: Verify properties a, b and c of the definition of a subspace.

a. The zero vector of $\mathbf{R}^3$ is in H (let $a =$ _______ and $b =$ _______).

b. Adding two vectors in H always produces another vector whose second entry is ______ and therefore the sum of two vectors in H is also in H. (H is closed under addition)

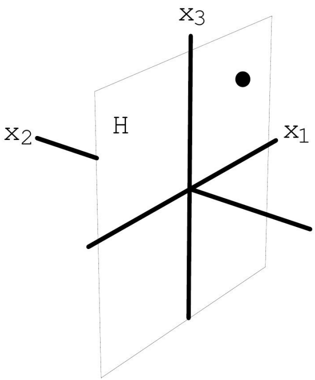
c. Multiplying a vector in H by a scalar produces another vector in H (H is closed under scalar multiplication).

Since properties a, b, and c hold, V is a subspace of $\mathbf{R}^3$.

Subspaces: Example (cont.)

Note

Vectors $(a, 0, b)$ in H look and act like the points (a, b) in $\mathbf{R}^2$.



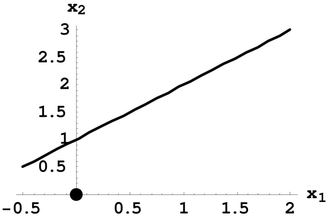
Graphical Depiction of H

Subspaces: Example

Example

Is $H = \left\{ \begin{bmatrix} x \\ x+1 \end{bmatrix} : x \text{ is real} \right\}$ a subspace of ______?
I.e., does H satisfy properties a, b and c?

Solution: For H *to be a subspace of* $\mathbf{R}^2$, *all three properties must hold*



Property (a) fails

Property (a) is not true because ________________________________.
Therefore H is not a subspace of $\mathbf{R}^2$.

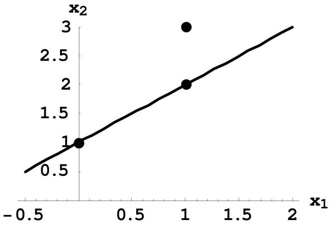
Subspaces: Example (cont.)

Another way to show that H is not a subspace of $\mathbf{R}^2$:

Let

$$\mathbf{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{ then } \mathbf{u}+\mathbf{v} = \begin{bmatrix} \quad\quad \end{bmatrix}$$

and so $\mathbf{u}+\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, which is ____ in H. So property (b) fails and so H is not a subspace of $\mathbf{R}^2$.



Property (b) fails

A Shortcut for Determining Subspaces

Theorem (1)

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in a vector space V, then Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

Proof: In order to verify this, check properties a, b and c of definition of a subspace.

a. $\mathbf{0}$ is in Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ since

$$\mathbf{0} = \_\_\_\_\_\mathbf{v}_1 + \_\_\_\_\_\mathbf{v}_2 + \cdots + \_\_\_\_\_\mathbf{v}_p$$

b. To show that Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ closed under vector addition, we choose two arbitrary vectors in Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$:

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p$$

and

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p.$$

A Shortcut for Determining Subspaces (cont.)

Then

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p) + (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p) \\ &= (\_\_\_\mathbf{v}_1 + \_\_\_\_\mathbf{v}_1) + (\_\_\_\_\mathbf{v}_2 + \_\_\_\_\mathbf{v}_2) + \cdots + (\_\_\_\_\mathbf{v}_p + \_\_\_\_\mathbf{v}_p) \\ &= (a_1+b_1)\,\mathbf{v}_1 + (a_2+b_2)\,\mathbf{v}_2 + \cdots + (a_p+b_p)\,\mathbf{v}_p.\end{aligned}$$

So $\mathbf{u}+\mathbf{v}$ is in Span$\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$.

c. To show that Span$\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$ closed under scalar multiplication, choose an arbitrary number c and an arbitrary vector in Span$\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$:

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p.$$

A Shortcut for Determining Subspaces (cont.)

Then

$$c\mathbf{v} = c\left(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p\right)$$

$$= \_\_\_\_\_\_\mathbf{v}_1 + \_\_\_\_\_\_\mathbf{v}_2 + \cdots + \_\_\_\_\_\_\mathbf{v}_p$$

So $c\mathbf{v}$ is in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Since properties a, b and c hold, $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

Determining Subspaces: Recap

Recap

1. To show that H is a subspace of a vector space, use Theorem 1.
2. To show that a set is not a subspace of a vector space, provide a specific example showing that at least one of the axioms a, b or c (from the definition of a subspace) is violated.

Determining Subspaces: Example

Example

Is $V = \{(a+2b, 2a-3b) : \ a \text{ and } b \text{ are real}\}$ a subspace of $\mathbf{R}^2$? Why or why not?

Solution: Write vectors in V in column form:

$$\begin{bmatrix} a+2b \\ 2a-3b \end{bmatrix} = \begin{bmatrix} a \\ 2a \end{bmatrix} + \begin{bmatrix} 2b \\ -3b \end{bmatrix}$$

$$= \_\_\_\_\_ \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \_\_\_\_\_ \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

So $V = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and therefore V is a subspace of _____ by Theorem 1.

Determining Subspaces: Example

Example

Is $H = \left\{ \begin{bmatrix} a+2b \\ a+1 \\ a \end{bmatrix} : a \text{ and } b \text{ are real} \right\}$ a subspace of $\mathbf{R}^3$?
Why or why not?

Solution: $\mathbf{0}$ is not in H since $a = b = 0$ or any other combination of values for a and b does not produce the zero vector. So property _____ fails to hold and therefore H is not a subspace of $\mathbf{R}^3$.

Determining Subspaces: Example

Example

Is the set H of all matrices of the form $\begin{bmatrix} 2a & b \\ 3a+b & 3b \end{bmatrix}$ a subspace of $M_{2\times 2}$? Explain.

Solution: Since

$$\begin{bmatrix} 2a & b \\ 3a+b & 3b \end{bmatrix} = \begin{bmatrix} 2a & 0 \\ 3a & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ b & 3b \end{bmatrix}$$

$$= a\begin{bmatrix} & \\ & \end{bmatrix} + b\begin{bmatrix} & \\ & \end{bmatrix}.$$

Therefore H =Span $\left\{ \begin{bmatrix} 2 & 0 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 3 \end{bmatrix} \right\}$ and so H is a subspace of $M_{2\times 2}$.

Linear Combinations and Span

Given two vectors $\mathbf{v}$ and $\mathbf{w}$, a **linear combination** of $\mathbf{v}$ and $\mathbf{w}$ is any vector of the form

$$a\mathbf{v} + b\mathbf{w}$$

where a and b are scalars. For example, the vector $(6, 8, 10)$ is a linear combination of the vectors $(1, 1, 1)$ and $(1, 2, 3)$, since

$$\begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

For example, a linear combination of three vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ would have the form $a\mathbf{u} + b\mathbf{v} + c\mathbf{w}$, where a, b, and c are scalars.

More generally, a **linear combination** of n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is any vector of the form

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

where $a_1, a_2, \ldots, a_n$ are scalars. For $n = 2$, this reduces to the definition for two vectors given above.

It is all right if some of the scalars in a linear combination are either zero or negative. For example, if $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors, then

$$2\mathbf{u} - 3\mathbf{v} + 4\mathbf{w}, \qquad 3\mathbf{u} + 5\mathbf{w}, \qquad \mathbf{v} + \mathbf{w}, \qquad \mathbf{w} - \mathbf{u}, \qquad \text{and} \qquad 5\mathbf{v}$$

are some possible linear combinations of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$.

We will sometimes want to discuss linear combinations of a single vector. If $\mathbf{v}$ is a vector, a linear combination of just $\mathbf{v}$ is the same thing as a scalar multiple of $\mathbf{v}$:

$$a\mathbf{v}.$$

Thus $(3, 12, 6)$ is a linear combination of $(1, 4, 2)$, since $(3, 12, 6) = 3(1, 4, 2)$.

Expressing a Vector as a Linear Combination

Sometimes you want to express one vector as a linear combination of others. For example, can we express the vector $(8, 3, 3)$ as a linear combination of $(1, 1, 1)$ and $(1, 0, 0)$? A moment's thought reveals the answer:

$$\begin{bmatrix} 8 \\ 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

For more complicated examples, you can express one vector as a linear combination of others by solving a system of linear equations.

EXAMPLE 1 Express the vector $(9, 6)$ as a linear combination of the vectors $(1, 2)$ and $(1, -4)$.

SOLUTION We are looking for scalars x_1 and x_2 so that

$$x_1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2\begin{bmatrix} 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \end{bmatrix}.$$

We can write this equation as a system of linear equations:

$$\begin{aligned} x_1 + x_2 &= 9 \\ 2x_1 - 4x_2 &= 6 \end{aligned}$$

Solving gives $x_1 = 7$ and $x_2 = 2$. Thus

$$\begin{bmatrix} 9 \\ 6 \end{bmatrix} = 7\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -4 \end{bmatrix}.$$

■

EXAMPLE 2 Determine whether the vector $(2, 1, 3)$ is a linear combination of the vectors $(1, 2, 3)$ and $(2, 3, 1)$.

SOLUTION We are looking for scalars x_1 and x_2 so that

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

We can write this equation as a system of linear equations:

$$\begin{aligned} x_1 + 2x_2 &= 2 \\ 2x_1 + 3x_2 &= 1 \\ 3x_1 + x_2 &= 3 \end{aligned}$$

which we can solve using row reduction:

$$\left[\begin{array}{cc|c} 1 & 2 & 2 \\ 2 & 3 & 1 \\ 3 & 1 & 3 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & -1 & -3 \\ 0 & -5 & -3 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 1 & 3 \\ 0 & -5 & -3 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 12 \end{array}\right].$$

If we had found a solution for x_1 and x_2, it would have meant that $(2, 1, 3)$ *was* a linear combination of $(1, 2, 3)$ and $(2, 3, 1)$.

By this point, it has become clear that the system of linear equations has no solutions. We conclude that $(2, 1, 3)$ is not a linear combination of $(1, 2, 3)$ and $(2, 3, 1)$. ■

The Span of Vectors

The **span** of a collection of vectors is the set of all possible linear combinations of them. For example, the span of the vectors $(1, 5, 3)$ and $(2, 1, 7)$ is the set of all vectors of the form

$$s \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix} + t \begin{bmatrix} 2 \\ 1 \\ 7 \end{bmatrix}$$

as s and t range over all possible scalars.

EXAMPLE 3 Describe the span of the vectors $(1, 0, 0)$ and $(0, 1, 1)$.

SOLUTION A linear combination of the vectors $(1, 0, 0)$ and $(0, 1, 1)$ has the form

$$s \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} s \\ t \\ t \end{bmatrix}$$

where s and t many be any real numbers. Hence, the span of the vectors $(1, 0, 0)$ and $(0, 1, 1)$ is the set of all vectors in $\mathbb{R}^3$ whose second and third entries are the same. ■

EXAMPLE 4 Describe the span of the vector $(1, 4)$.

SOLUTION Recall that a linear combination of $(1, 4)$ is just any scalar multiple of $(1, 4)$:

$$t \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} t \\ 4t \end{bmatrix}.$$

As t ranges over all real numbers, this gives all possible vectors whose y-component is 4 times the x-component. Thus, the span of the vector $(1, 4)$ is the line $y = 4x$ in $\mathbb{R}^2$. ■

There is an exception to this rule. If $\mathbf{v}$ is the zero vector, then any multiple of $\mathbf{v}$ is again the zero vector, so the span of $\mathbf{v}$ is not a line.

In general, if $\mathbf{v}$ is any vector in $\mathbb{R}^n$, then the span of $\mathbf{v}$ is the line in $\mathbb{R}^n$ consisting of all scalar multiples of $\mathbf{v}$. That is, the span of $\mathbf{v}$ is the line in $\mathbb{R}^n$ that goes through the origin as well as the point $\mathbf{v}$.

Something similar happens if you take the span of two vectors: the result is usually a plane. In particular:

- If you take the span of two vectors in $\mathbb{R}^2$, the result is usually the entire plane $\mathbb{R}^2$.
- If you take the span of two vectors in $\mathbb{R}^3$, the result is usually a plane through the origin in 3-dimensional space.
- Similarly, if you take the span of two vectors in $\mathbb{R}^n$ (where $n > 3$), the result is usually a plane through the origin in n-dimensional space.

More precisely, if you take the span of two vectors $\mathbf{v}$ and $\mathbf{w}$, the result is the plane that goes through the origin as well as the points $\mathbf{v}$ and $\mathbf{w}$.

The word "usually" is important here. For example, if $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^3$ and $\mathbf{w} = (0,0,0)$, then the span of $\mathbf{v}$ and $\mathbf{w}$ will be the same as all the multiples of $\mathbf{v}$, which is just a line. More generally, if $\mathbf{w}$ is itself a multiple of $\mathbf{v}$, then every linear combination of $\mathbf{v}$ and $\mathbf{w}$ is again a multiple of $\mathbf{v}$, so the span of $\mathbf{v}$ and $\mathbf{w}$ is just a line. This makes sense geometrically: if $\mathbf{w}$ is a multiple of $\mathbf{v}$ then $\mathbf{v}$ and $\mathbf{w}$ line on the same line through the origin, and the span of $\mathbf{v}$ and $\mathbf{w}$ is this line.

Using this intuition, it's not hard to find vectors whose span is a given line or plane.

EXAMPLE 5 Find a vector in $\mathbb{R}^2$ whose span is the line $y = 2x$.

SOLUTION We just need any vector at all that lies on this line, other than the zero vector. For example, the span of the vector $(1,2)$ is the line $y = 2x$. ■

EXAMPLE 6 Find two vectors in $\mathbb{R}^3$ whose span is the plane $2x - 6y + 5z = 0$.

SOLUTION Again, any two vectors on this plane will work, as long as they are not multiples of each other. But how can we make up points on this plane?

The simplest method is to choose values for two of the variables, and then solve for the third variable. That is, we think of the plane as

It really does work to just make up points on the plane. For example, observe that $(3,1,0)$ and $(5,0,-2)$ both lie on the plane $2x - 6y + 5z = 0$. These are not multiples of one another, so their span is the given plane.

$$x = 3y - \frac{5}{2}z$$

where y and z are free variables. Setting $y = 1$ and $z = 0$ gives the vector $(3,1,0)$, while setting $y = 0$ and $z = 1$ gives the vector $(-5/2, 0, 1)$. Both of these vectors lie on the plane, and they are not multiples of one another, so their span is the entire plane. ■

EXAMPLE 7 Let P be the plane in $\mathbb{R}^3$ that goes through the origin as well as the points $(5,1,3)$ and $(2,-1,2)$. Does the point $(4,5,0)$ lie on P?

SOLUTION The plane P is just the span of the vectors $(5,1,3)$ and $(2,-1,2)$. Thus the point $(4,5,0)$ will lie on this plane if and only if it can be expressed as a linear combination of $(5,1,3)$ and $(2,-1,2)$. So we want to know whether the equation

$$x_1 \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

has any solutions for x_1 and x_2. To find out, we must solve the linear system

$$\begin{aligned} 5x_1 + 2x_2 &= 4 \\ x_1 - x_2 &= 5 \\ 3x_1 + 2x_2 &= 0 \end{aligned}$$

Solving yields $x_1 = 2$ and $x_2 = -3$, i.e.

$$2\begin{bmatrix}5\\1\\3\end{bmatrix} - 3\begin{bmatrix}2\\-1\\2\end{bmatrix} = \begin{bmatrix}4\\5\\0\end{bmatrix}.$$

Since $(4,5,0)$ is a linear combination of $(5,1,3)$ and $(2,-1,2)$, this point does lie on the plane P. ■

EXERCISES

1–8 ■ Express the vector $\mathbf{w}$ as a linear combination of the given vectors $\mathbf{v}_i$. (Note: You ought to be able to solve problems 1–8 in your head.)

1. $\mathbf{v}_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}3\\5\end{bmatrix}$

2. $\mathbf{v}_1 = \begin{bmatrix}2\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\3\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}8\\9\end{bmatrix}$

3. $\mathbf{v}_1 = \begin{bmatrix}2\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}1\\0\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}13\\4\end{bmatrix}$

4. $\mathbf{v}_1 = \begin{bmatrix}5\\7\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}1\\4\end{bmatrix}$ $\mathbf{w} = \begin{bmatrix}5\\20\end{bmatrix}$

5. $\mathbf{v}_1 = \begin{bmatrix}-2\\3\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}-8\\12\end{bmatrix}$

6. $\mathbf{v}_1 = \begin{bmatrix}2\\0\\5\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\2\\0\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}4\\-6\\10\end{bmatrix}$

7. $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}2\\3\\8\end{bmatrix}$

8. $\mathbf{v}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\2\\0\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}0\\0\\3\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}5\\9\\2\end{bmatrix}$

9–12 ■ Use the method of Example 1 to express the vector $\mathbf{w}$ as a linear combination of the vectors $\mathbf{v}_i$.

9. $\mathbf{v}_1 = \begin{bmatrix}3\\7\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}6\\-5\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}21\\30\end{bmatrix}$

10. $\mathbf{v}_1 = \begin{bmatrix}1\\2\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}3\\4\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}-2\\1\end{bmatrix}$

11. $\mathbf{v}_1 = \begin{bmatrix}1\\3\\7\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}2\\7\\4\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}2\\6\\5\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}3\\7\\23\end{bmatrix}$

12. $\mathbf{v}_1 = \begin{bmatrix}1\\2\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}4\\3\\2\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}8\\1\\2\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}-3\\4\\1\end{bmatrix}$

13–14 ■ Determine whether the vector $\mathbf{w}$ lies in the span of the vectors $\mathbf{v}_i$.

13. $\mathbf{v}_1 = \begin{bmatrix}1\\2\\5\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}3\\1\\4\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}14\\3\\15\end{bmatrix}$

14. $\mathbf{v}_1 = \begin{bmatrix}1\\3\\7\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}2\\1\\4\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}4\\7\\3\end{bmatrix}$

15. Let L be the line in $\mathbb{R}^3$ that goes through the points $(0,0,0)$ and $(1,4,2)$. Does the point $(3,12,5)$ lie on this line? Explain.

16. Let P be the plane in $\mathbb{R}^3$ that goes through the points $(0,0,0)$, $(1,2,0)$, and $(0,0,1)$. Does the point $(3,6,4)$ lie on this plane? Explain.

17–22 ■ Determine whether the span of the given vectors $\mathbf{v}_i$ is a single point, a line, or all of $\mathbb{R}^2$. If the span is a line, give the equation for the line.

17. $\mathbf{v}_1 = \begin{bmatrix}3\\6\end{bmatrix}$.

18. $\mathbf{v}_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\end{bmatrix}$.

19. $\mathbf{v}_1 = \begin{bmatrix}0\\0\end{bmatrix}$.

20. $\mathbf{v}_1 = \begin{bmatrix}2\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}-1\\5\end{bmatrix}$.

21. $\mathbf{v}_1 = \begin{bmatrix}2\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\0\end{bmatrix}$.

22. $\mathbf{v}_1 = \begin{bmatrix}1\\3\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}2\\6\end{bmatrix}$.

23. Find a vector $\mathbf{v}_1$ in $\mathbb{R}^2$ whose span is the line $2x + 3y = 0$.

24. Find two vectors $\mathbf{v}_1, \mathbf{v}_2$ in $\mathbb{R}^3$ whose span is the plane $x + 3y - 4z = 0$.

Answers

1. $\mathbf{w} = 3\mathbf{v}_1 + 5\mathbf{v}_2$ **2.** $\mathbf{w} = 4\mathbf{v}_1 + 3\mathbf{v}_2$ **3.** $\mathbf{w} = 4\mathbf{v}_1 + 5\mathbf{v}_2$ **4.** $\mathbf{w} = 0\mathbf{v}_1 + 5\mathbf{v}_2$ (or simply $\mathbf{w} = 5\mathbf{v}_2$)

5. $\mathbf{w} = 4\mathbf{v}_1$ **6.** $\mathbf{w} = 2\mathbf{v}_1 - 3\mathbf{v}_2$ **7.** $\mathbf{w} = 2\mathbf{v}_1 + 3\mathbf{v}_2 + 8\mathbf{v}_3$ **8.** $\mathbf{w} = 5\mathbf{v}_1 + 2\mathbf{v}_2 - \mathbf{v}_3$

9. $\mathbf{w} = 5\mathbf{v}_1 + \mathbf{v}_2$ **10.** $\mathbf{w} = \frac{11}{2}\mathbf{v}_1 - \frac{5}{2}\mathbf{v}_2$ **11.** $\mathbf{w} = 3\mathbf{v}_1 - 2\mathbf{v}_2 + 2\mathbf{v}_3$

12. Many answers are possible. For example, $\mathbf{w} = 5\mathbf{v}_1 - 2\mathbf{v}_2 + 0\mathbf{v}_3$. **13.** Yes, since $\mathbf{w} = -\mathbf{v}_1 + 5\mathbf{v}_2$.

14. No, since $\mathbf{w}$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. **15.** No, since $(3,12,5)$ is not a scalar multiple of $(1,4,2)$.

16. Yes, since $(3,6,4)$ is a linear combination of $(1,2,0)$ and $(0,0,1)$. **17.** The span is the line $y = 2x$.

18. The span is all of $\mathbb{R}^2$ **19.** The span is the single point $(0,0)$. **20.** The span is all of $\mathbb{R}^2$.

21. The span is the line $y = \frac{1}{2}x$. **22.** The span is the line $y = 3x$.

23. Many answers are possible. For example, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ suffices.

24. Many answers are possible. For example, $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$ suffice.

Last lecture: Examples and the column space of a matrix
Suppose that A is an $n \times m$ matrix.

Definition The column space of A is the vector subspace $\mathrm{Col}(A)$ of $\mathbb{R}^n$ which is spanned by the columns of A.

That is, if $A = [a_1, a_2, \ldots, a_m]$ then $\mathrm{Col}(A) = \mathrm{Span}\,(a_1, a_2, \ldots, a_m)$.

Linear dependence and independence (chapter. 4)

- If V is *any* vector space then $V = \mathrm{Span}(V)$.
- Clearly, we can find smaller sets of vectors which span V.
- This lecture we will use the notions of linear independence and linear dependence to find the smallest sets of vectors which span V.
- It turns out that there are many "smallest sets" of vectors which span V, and that the number of vectors in these sets is always the same.

  This number is the dimension of V.

Linear dependence—motivation Let lecture we saw that the two sets of vectors $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \\ 13 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}$ do not span $\mathbb{R}^3$.

- The problem is that

  $\begin{bmatrix} 5 \\ 9 \\ 13 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$ and

  $\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$.

- Therefore,

  $$\mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \\ 13 \end{bmatrix} \right) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} \right)$$

  and

  $$\mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} \right).$$

- Notice that we can rewrite the two equations above in the following form:

  $2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} - \begin{bmatrix} 5 \\ 9 \\ 13 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and

  $3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

This is the key observation about spanning sets.

Definition

Suppose that V is a vector space and that $x_1, x_2, \ldots, x_k$ are vectors in V.

The set of vectors $\{x_1, x_2, \ldots, x_k\}$ is linearly dependent

if

$$r_1 x_1 + r_2 x_2 + \cdots + r_k x_k = 0$$

for some $r_1, r_2, \ldots, r_k \in \mathbb{R}$ where at least one of $r_1, r_2, \ldots, r_k$ is non–zero.

Example

$2\begin{bmatrix}1\\2\\3\end{bmatrix} + \begin{bmatrix}3\\5\\7\end{bmatrix} - \begin{bmatrix}5\\9\\13\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$ and

$3\begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}3\\5\\7\end{bmatrix} - \begin{bmatrix}0\\1\\2\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$

So the two sets of vectors $\left\{\begin{bmatrix}5\\9\\13\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}3\\5\\7\end{bmatrix}\right\}$ and $\left\{\begin{bmatrix}0\\1\\2\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}3\\5\\7\end{bmatrix}\right\}$ are linearly dependent.

Question Suppose that $x, y \in V$. When are x and y linearly dependent?

Question What do linearly dependent vectors look like in $\mathbb{R}^2$ and $\mathbb{R}^3$?

Example

Let $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $y = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ and $z = \begin{bmatrix} 0 \\ 4 \\ 8 \end{bmatrix}$. Is $\{x_1, x_2, x_3\}$ linearly dependent?

We have to determine whether or not we can find real numbers r, s, t, which are not all zero, such that $rx + sy + tz = 0$.

To find all possible r, s, t we have to solve the augmented matrix equation:

$$\left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 2 & 2 & 4 & 0 \\ 3 & 1 & 8 & 0 \end{array}\right] \xrightarrow[R_3:=R_3-3R_1]{R_2:=R_2-2R_1} \left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 0 & -4 & 4 & 0 \\ 0 & -8 & 8 & 0 \end{array}\right]$$

$$\xrightarrow{R_3:=R_3-2R_2} \left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

So this set of equations has a non–zero solution.
Therefore, $\{x, y, z\}$ is a linearly dependent set of vectors.

To be explicit, $3\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Linear dependence–Example II

Example Consider the polynomials $p(x) = 1+3x+2x^2$, $q(x) = 3 + x + 2x^2$ and $r(x) = 2x + x^2$ in $\mathbb{P}_2$.
Is $\{p(x), q(x), r(x)\}$ linearly dependent?

We have to decide whether we can find real numbers r, s, t, which are not all zero, such that
$rp(x) + sq(x) + tr(x) = 0$.
That is:

$$\begin{aligned} 0 &= r(1 + 3x + 2x^2) + s(3 + x + 2x^2) + t(2x + x^2) \\ &= (r+3s)+(3r+s+2t)x+(2r+2s+t)x^2. \end{aligned}$$

This corresponds to solving the following system of linear equations

$$\begin{array}{cccc} r & +3s & & = 0 \\ 3r & +s & +2t & = 0 \\ 2r & +2s & +t & = 0 \end{array}$$

We compute:

$$\begin{bmatrix} 1 & 3 & 0 \\ 3 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \xrightarrow[R_3:=R_3-2R_1]{R_2:=R_2-3R_1} \begin{bmatrix} 1 & 3 & 0 \\ 0 & -8 & 2 \\ 0 & -4 & 1 \end{bmatrix}$$

$$\xrightarrow{R_2:=R_2-R_3} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & -4 & 1 \end{bmatrix}$$

Hence, $\{p(x), q(x), r(x)\}$ is linearly dependent.

Linear independence

In fact, we do not care so much about linear dependence as about its *opposite* linear independence:

Definition
Suppose that V is a vector space.

The set of vectors $\{x_1, x_2, \ldots, x_k\}$ in V is linearly independent if the only scalars $r_1, r_2, \ldots, r_k \in \mathbb{R}$ such that

$$r_1 x_1 + r_2 x_2 + \cdots + r_k x_k = 0$$

are $r_1 = r_2 = \cdots = r_k = 0$.

(That is, $\{x_1, \ldots, x_k\}$ is not linearly dependent!)

- If $\{x_1, x_2, \ldots, x_k\}$ are linearly independent then it is not possible to write any of these vectors as a linear combination of the remaining vectors.

For example, if $x_1 = r_2 x_2 + r_3 x_3 + \cdots + r_k x_k$ then

$$-x_1 + r_2 x_2 + r_3 x_3 + \cdots + r_k x_k = 0$$

$\Longrightarrow$ all of these coefficients must be zero!!??!!

Linear independence—examples

The following sets of vectors are all linearly independent:

- $\left\{[1]\right\}$ is a linearly independent subset of $\mathbb{R}$.
- $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is a linearly independent subset of $\mathbb{R}^2$.
- $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$ is a linearly independent subset of $\mathbb{R}^3$.
- $\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}$ is a linearly independent subset of $\mathbb{R}^4$.
- $\left\{\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \dots, \begin{bmatrix}0\\\vdots\\1\\0\end{bmatrix}, \begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix}\right\}$ is a linearly independent subset of $\mathbb{R}^m$.
- $\{1\}$ is a linearly independent subset of $\mathbb{P}_0$.
- $\{1, x\}$ is a linearly independent subset of $\mathbb{P}_1$.
- $\{1, x, x^2\}$ is a linearly independent subset of $\mathbb{P}_2$.

- $\{1, x, x^2, \ldots, x^n\}$ is a linearly independent subset of $\mathbb{P}_n$.

Linear independence—example 2

Example

Let $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $y = \begin{bmatrix} 3 \\ 2 \\ 9 \end{bmatrix}$ and $z = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$.

Is the set $\{x_1, x_2, x_3\}$ linearly independent?

We have to determine whether or not we can find real numbers r, s, t, which are not all zero, such that $rx + sy + tz = 0$.

Once again, to find all possible r, s, t we have to solve the augmented matrix equation:

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 9 & -1 & 0 \end{array}\right] \xrightarrow[R_3 := R_3 - 3R_1]{R_2 := R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 0 & -4 & -8 & 0 \\ 0 & 0 & -16 & 0 \end{array}\right]$$

$$\xrightarrow[R_3 := -\frac{1}{16} R_3]{R_2 := -\frac{1}{4} R_2} \left[\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Hence, $rx + sy + tz = 0$ only if $r = s = t = 0$.

Therefore, $\{x_1, x_2, x_3\}$ is a linearly independent subset of $\mathbb{R}^3$.

Linear independence—example 3

Example

Let $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $x_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}$ and $x_4 = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 7 \end{bmatrix}$.

Is $\{x_1, x_2, x_3, x_4\}$ linear dependent or linearly independent?

Again, we have to solve the corresponding system of linear equations:

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 1 & 5 \\ 1 & 4 & 2 & 7 \end{bmatrix} \xrightarrow[R_4 = R_4 - R_1]{\substack{R_2 = R_2 - R_1 \\ R_3 = R_3 - R_1}} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 0 & 2 \\ 0 & 3 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow[R_4 = R_4 - 3R_2]{R_3 = R_3 - 2R_2} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & -2 & -2 \end{bmatrix}$$

$$\xrightarrow{R_4 = R_4 - R_3} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence, after much work, we see that $\{x_1, x_2, x_3, x_4\}$ is linearly dependent.

Linear independence—example 4

Example

Let $X = \{\sin x, \cos x\} \subset \mathbb{F}$.

Is X linearly dependent or linearly independent?

Suppose that $s \sin x + t \cos x = 0$.

Notice that this equation holds for all $x \in \mathbb{R}$, so

$$\begin{aligned} x = 0: &\qquad s \cdot 0 + t \cdot 1 &= 0 \\ x = \tfrac{\pi}{2}: &\qquad s \cdot 1 + t \cdot 0 &= 0 \end{aligned}$$

Therefore, we must have $s = 0 = t$.

Hence, $\{\sin x, \cos x\}$ is linearly independent.

What happens if we tweak this example by a little bit?

Example Is $\{\cos x, \sin x, x\}$ is linearly independent?

If $s \cos x + t \sin x + r = 0$ then

$$\begin{aligned} x = 0: &\qquad s \cdot 0 + t \cdot 1 + r \cdot 0 &= 0 \\ x = \tfrac{\pi}{2}: &\qquad s \cdot 1 + t \cdot 0 + r \cdot \tfrac{\pi}{2} &= 0 \\ x = \tfrac{\pi}{4}: &\qquad s \cdot \tfrac{1}{\sqrt{2}} + t \cdot \tfrac{1}{\sqrt{2}} + r \cdot \tfrac{\pi}{4} &= 0 \end{aligned}$$

Therefore, $\{\cos x, \sin x, x\}$ is linearly independent.

Linear independence—last example

Example

Show that $X = \{e^x, e^{2x}, e^{3x}\}$ is a linearly independent subset of $\mathbb{F}$.

Suppose that $re^x + se^{2x} + te^{3x} = 0$.

Then:

$$\begin{array}{lll} x = 0 & r + s + t & = 0, \\ x = 1 & re + se^2 + te^3 & = 0, \\ x = 2 & re^2 + se^4 + te^6 & = 0, \end{array}$$

So we have to solve the matrix equation:

$$\begin{bmatrix} 1 & 1 & 1 \\ e & e^2 & e^3 \\ e^2 & e^4 & e^6 \end{bmatrix} \xrightarrow[R_3 := \frac{1}{e^2} R_3]{R_2 := \frac{1}{e} R_2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e & e^2 \\ 1 & e^2 & e^4 \end{bmatrix}$$

$$\xrightarrow[R_3 := R_3 - R_1]{R_2 := R_2 - R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & e-1 & e^2-1 \\ 0 & e^2-1 & e^4-1 \end{bmatrix}$$

$$\xrightarrow[R_3 := \frac{1}{e^2-1} R_3]{R_2 := \frac{1}{e-1} R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & e+1 \\ 0 & 1 & e^2+1 \end{bmatrix}$$

$$\xrightarrow{R_3 := R_3 - R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & e+1 \\ 0 & 0 & e^2-e \end{bmatrix}$$

Therefore, $\{e^x, e^{2x}, e^{3x}\}$ is a set of linearly independent functions in the vector space $\mathbb{F}$.

The Basis of a Vector Space:
We now combine the ideas of spanning sets and linear independence.

Definition Suppose that V is a vector space.
A basis of V is a set of vectors $\{x_1, x_2, \ldots, x_k\}$ in V such that

- $V = \text{Span}(x_1, x_2, \ldots, x_k)$ and
- $\{x_1, x_2, \ldots, x_k\}$ is linearly independent.

Examples

- $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right.$ is a basis of $\mathbb{R}^2$.
- $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of $\mathbb{R}^3$.
- $\left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of $\mathbb{R}^m$.
- $\{1, x, x^2\}$ is a basis of $\mathbb{P}_2$.
- $\{1, x, x^2, \ldots, x^n\}$ is a basis of $\mathbb{P}_n$.
- In general, if W is a vector subspace of V then the challenge is to find a basis for W.

4.5 Basis and Dimension of a Vector Space

In the section on spanning sets and linear independence, we were trying to understand what the elements of a vector space looked like by studying how they could be generated. We learned that some subsets of a vector space could generate the entire vector space. Such subsets were called spanning sets. Other subsets did not generate the entire space, but their span was still a subspace of the underlying vector space. In some cases, the number of vectors in such a set was redundant in the sense that one or more of the vectors could be removed, without changing the span of the set. In other cases, there was not a unique way to generate some vectors in the space. In this section, we want to make this process of generating all the elements of a vector space more reliable, more efficient.

4.5.1 Basis of a Vector Space

Definition 297 *Let V denote a vector space and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ a subset of V. S is called a* ***basis*** *for V if the following is true:*

1. *S spans V.*
2. *S is linearly independent.*

This definition tells us that a basis has to contain enough vectors to generate the entire vector space. But it does not contain too many. In other words, if we removed one of the vectors, it would no longer generate the space.

A basis is the vector space generalization of a coordinate system in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Example 298 *We have already seen that the set $S = \{\mathbf{e}_1, \mathbf{e}_2\}$ where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ was a spanning set of $\mathbb{R}^2$. It is also linearly independent for the only solution of the vector equation $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 = 0$ is the trivial solution. Therefore, S is a basis for $\mathbb{R}^2$. It is called the* ***standard basis*** *for $\mathbb{R}^2$. These vectors also have a special name. $(1, 0)$ is $\mathbf{i}$ and $(0, 1)$ is $\mathbf{j}$.*

Example 299 *Similarly, the standard basis for $\mathbb{R}^3$ is the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$ and $\mathbf{e}_3 = (0, 0, 1)$. These vectors also have a special name. They are $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ respectively.*

Example 300 *Prove that $S = \left\{1, x, x^2\right\}$ is a basis for P_2, the set of polynomials of degree less than or equal to 2.*
We need to prove that S spans P_2 and is linearly independent.

- *S spans P_2. We already did this in the section on spanning sets. A typical polynomial of degree less than or equal to 2 is $ax^2 + bx + c$.*
- *S is linearly independent. Here, we need to show that the only solution to $a(1) + bx + cx^2 = 0$ (where 0 is the zero polynomial) is $a = b = c = 0$.*

> *From algebra, we remember that two polynomials are equal if and only if their corresponding coefficients are equal. The zero polynomial has all its coefficients equal to zero. So,* $a(1) + bx + cx^2 = 0$ *if and only if* $a = 0$, $b = 0$, $c = 0$. *Which proves that* S *is linearly independent.*

We will see more examples shortly.

The next theorem outlines an important difference between a basis and a spanning set. Any vector in a vector space can be represented in a unique way as a linear combination of the vectors of a basis..

Theorem 301 *Let* V *denote a vector space and* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *a basis of* V. *Every vector in* V *can be written in a unique way as a linear combination of vectors in* S.

Proof. *Since* S *is a basis, we know that it spans* V. *If* $\mathbf{v} \in V$, *then there exists scalars* $c_1, c_2, ..., c_n$ *such that* $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_n\mathbf{u}_n$. *Suppose there is another way to write* $\mathbf{v}$. *That is, there exist scalars* $d_1, d_2, ..., d_n$ *such that* $\mathbf{v} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + ... + d_n\mathbf{u}_n$. *Then,* $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_n\mathbf{u}_n = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + ... + d_n\mathbf{u}_n$. *In other words,* $(c_1 - d_1)\mathbf{u}_1 + (c_2 - d_2)\mathbf{u}_2 + ... + (c_n - d_n)\mathbf{u}_n = 0$. *Since* S *is a basis, it must be linearly independent. The unique solution to* $(c_1 - d_1)\mathbf{u}_1 + (c_2 - d_2)\mathbf{u}_2 + ... + (c_n - d_n)\mathbf{u}_n = 0$ *must be the trivial solution. It follows that* $c_i - d_i = 0$ *for* $i = 1, 2, ..., n$ *in other words* $c_i = d_i$ *for* $i = 1, 2, ..., n$. *Therefore, the two representations of* $\mathbf{v}$ *are the same.* ■

Remark 302 *We say that any vector* $\mathbf{v}$ *of* V *has a unique representation with respect to the basis* S. *The scalars used in the linear representation are called the coordinates of the vector. For example, the vector* (x, y) *can be represented in the basis* $\{(1, 0), (0, 1)\}$ *by the linear combination* $(x, y) = x(1, 0) + y(0, 1)$. *Thus,* x *and* y *are the coordinates of this vector (we knew that!).*

Definition 303 *If* V *is a vector space, and* $B = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *is an ordered basis of* V, *then we know that every vector* $\mathbf{v}$ *of* V *can be expressed as a linear combination of the vectors in* S *in a unique way. In others words, there exists unique scalars* $c_1, c_2, ..., c_n$ *such that* $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_n\mathbf{u}_n$. *These scalars are called the* ***coordinates*** *of* v *relative to the ordered basis* B.

Remark 304 *The term "ordered basis" simply means that the order in which we list the elements is important. Indeed it is since each coordinate is with respect to one of the vector in the basis. We know that in* $\mathbb{R}^2$, $(2, 3)$ *is not the same as* $(3, 2)$.

Example 305 *What are the coordinates of* $(1, 2, 3)$ *with respect to the basis* $\{(1, 1, 0), (0, 1, 1), (1, 1, 1)\}$*?*
One can indeed verify that this set is a basis for $\mathbb{R}^3$. *Finding the coordinates of* $(1, 2, 3)$ *with respect to this new basis amounts to finding the numbers* (a, b, c) *such that* $a(1, 1, 0) + b(0, 1, 1) + c(1, 1, 1) = (1, 2, 3)$. *This amounts to solving*

the system $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. *The solution is*

$$\begin{aligned} \begin{bmatrix} a \\ b \\ c \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \end{aligned}$$

The next theorem, deals with the number of vectors the basis of a given vector space can have. We will state the theorem without proof.

Theorem 306 *Let V denote a vector space and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ a basis of V.*

1. *Any subset of V containing more than n vectors must be dependent.*
2. *Any subset of V containing less than n vectors cannot span V.*

Proof. *We prove each part separately.*

1. *Consider $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ a subset of V where $r > n$. We must show that W in dependent. Since S is a basis, we can writ each $\mathbf{v}_i$ in term of elements in S. More specifically, there exists constants c_{ij} with $1 \leq i \leq r$ and $1 \leq j \leq n$ such that $\mathbf{v}_i = c_{i1}\mathbf{u}_1 + c_{i2}\mathbf{u}_2 + ... + c_{in}\mathbf{u}_n$. Consider the linear combination*

$$\sum_{j=1}^{r} d_j \mathbf{v}_j = \sum_{j=1}^{r} d_j \left(c_{j1}\mathbf{u}_1 + c_{j2}\mathbf{u}_2 + ... + c_{jn}\mathbf{u}_n \right) = 0$$

So, we must solve $\begin{cases} d_1c_{11} + d_1c_{12} + ... + d_1c_{1n} = 0 \\ d_2c_{21} + d_2c_{22} + ... + d_2c_{2n} = 0 \\ \vdots \\ d_rc_{r1} + d_rc_{r2} + ... + d_rc_{rn} = 0 \end{cases}$ *where the unknowns are $d_1, d_2, ..., d_r$. Since we have more unknowns than equations, we are guaranteed that this homogeneous system will have a nontrivial solution. Thus W is dependent.*

2. *Consider $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ a subset of V where $r < n$. We must show that W does not span V. We do it by contradiction. We assume it does span V and show this would imply that S is dependent. Suppose that there exists constants c_{ij} with $1 \leq i \leq n$ and $1 \leq j \leq r$ such that $\mathbf{u}_i = c_{i1}\mathbf{v}_1 + c_{i2}\mathbf{v}_2 + ... + c_{ir}\mathbf{v}_r$. Consider the linear combination*

$$\sum_{j=1}^{n} d_j \mathbf{u}_j = \sum_{j=1}^{r} d_j \left(c_{j1}\mathbf{v}_1 + c_{j2}\mathbf{v}_2 + ... + c_{jr}\mathbf{v}_r \right) = 0$$

So, we must solve $\begin{cases} d_1c_{11} + d_1c_{12} + ... + d_1c_{1r} = 0 \\ d_2c_{21} + d_2c_{22} + ... + d_2c_{2r} = 0 \\ \vdots \\ d_nc_{n1} + d_nc_{n2} + ... + d_nc_{nr} = 0 \end{cases}$ *where the unknowns are* $d_1, d_2, ..., d_n$. *Since we have more unknowns than equations, we are guaranteed that this homogeneous system will have a nontrivial solution. Thus* S *would be dependent. But it can't be since it is a basis.*

■

Corollary 307 *Let* V *denote a vector space. If* V *has a basis with* n *elements, then all the bases of* V *will have* n *elements.*
Proof. *Assume that* S_1 *is a basis of* V *with* n *elements and* S_2 *is another basis with* m *elements. We need to show that* $m = n$. *Since* S_1 *is a basis,* S_2 *being also a basis implies that* $m \leq n$. *If we had* $m > n$, *by the theorem,* S_2 *would be dependent, hence not a basis. Similarly, since* S_2 *is a basis,* S_1 *being also a basis implies that* $n \leq m$. *The only way we can have* $m \leq n$ *and* $n \leq m$ *is if* $m = n$. ■

4.5.2 Dimension of a Vector Space

All the bases of a vector space must have the same number of elements. This common number of elements has a name.

Definition 308 *Let* V *denote a vector space. Suppose a basis of* V *has* n *vectors (therefore all bases will have* n *vectors).* n *is called the* ***dimension*** *of* V. *We write* $\dim(V) = n$.

Remark 309 n *can be any integer.*

Definition 310 *A vector space* V *is said to be* ***finite-dimensional*** *if there exists a finite subset of* V *which is a basis of* V. *If no such finite subset exists, then* V *is said to be* ***infinite-dimensional****.*

Example 311 *We have seen, and will see more examples of finite-dimensional vector spaces. Some examples of infinite-dimensional vector spaces include* $F(-\infty, \infty)$, $C(-\infty, \infty)$, $C^m(-\infty, \infty)$.

Remark 312 *If* V *is just the vector space consisting of* $\{0\}$, *then we say that* $\dim(V) = 0$.

It is very important, when working with a vector space, to know whether its dimension is finite or infinite. Many nice things happen when the dimension is finite. The next theorem is such an example.

Theorem 313 *Let* V *denote a vector space such that* $\dim(V) = n < \infty$. *Let* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *be a subset of* V.

1. *If S spans V, then S is also linearly independent hence a basis for V.*
2. *If S is linearly independent, then S also spans V hence is a basis for V.*

This theorem says that in a finite dimensional space, for a set with as many elements as the dimension of the space to be a basis, it is enough if one of the two conditions for being a basis is satisfied.

4.5.3 Examples

Standard Basis of Known Spaces and Their Dimension

We look at some of the better known vector spaces under the standard operations, their standard bases, and their dimension.

Example 314 $\mathbb{R}^2$*, the set of all ordered pairs (x, y) where x and y are in $\mathbb{R}$. We have already seen that the standard basis for $\mathbb{R}^2$was $\{(1,0), (0,1)\}$. This basis has 2 elements, therefore,* $\dim\left(\mathbb{R}^2\right) = 2$.

Example 315 $\mathbb{R}^3$*, the set of all ordered triples (x, y, z) where x, y and z are in $\mathbb{R}$. Similarly, the standard basis for $\mathbb{R}^3$ is $\{(1,0,0), (0,1,0), (0,0,1)\}$. This basis has 3 elements, therefore,* $\dim\left(\mathbb{R}^3\right) = 3$.

Example 316 $\mathbb{R}^n$*, the set of all ordered n-tuples $(x_1, x_2, ..., x_n)$ where $x_1, x_2, ..., x_n$ are in $\mathbb{R}$. Similarly, the standard basis for $\mathbb{R}^n$ is $\{(1,0,0,...,0), (0,1,0,...,0), ..., (0,0,...,0,1)\}$. This basis has n elements, therefore* $\dim(\mathbb{R}^n) = n$.

Example 317 P_2*, the set of polynomials of degree less than or equal to 2. We have already seen that the standard basis for P_2 was $\left\{1, x, x^2\right\}$. This basis has 3 elements, therefore* $\dim(P_2) = 3$.

Example 318 P_3*, the set of polynomials of degree less than or equal to 3. Similarly, the standard basis for P_3 is $\left\{1, x, x^2, x^3\right\}$. This basis has 4 elements, therefore* $\dim(P_3) = 4$.

Example 319 P_n*, the set of polynomials of degree less than or equal to n. Similarly, the standard basis for P_n is $\left\{1, x, x^2, ..., x^n\right\}$. This basis has $n+1$ elements, therefore* $\dim(P_n) = n + 1$.

Example 320 $M_{3,2}$*, the set of 2×3 matrices. The user will check as an exercise that a basis for $M_{3,2}$ is* $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. *This is the standard basis for $M_{3,2}$. This basis has 6 elements, therefore* $\dim(M_{3,2}) = 6$.

Example 321 $M_{m,n}$*, the set of $m \times n$ matrices. The standard basis for $M_{m,n}$ is the set $\{B_{ij} | i = 1, 2, ..., m$ and $j = 1, 2, ..., n\}$ where B_{ij} is the $m \times n$ matrix whose entries are all zeros, except for the ij entry which is 1. This set has mn elements, therefore* $\dim(M_{m,n}) = mn$

Example 322 *P, the set of all polynomials. This space does not have a finite dimension, in other words* $\dim(P) = \infty$. *To see this, assume that it has finite dimension, say* $\dim(P) = n$. *Let S be a basis for P. S has n elements. Let m be the highest degree of the polynomials which appear in S. Then, x^{m+1} cannot be obtained by linear combination of elements of S which contradicts the fact that S is a basis.*

Example 323 *Show that $B = \{(1,1), (0,1\}$ is a basis for $\mathbb{R}^2$.*
We know that $\dim(\mathbb{R}^2) = 2$. *Since B has two elements, it is enough to show that B is either independent or that it spans $\mathbb{R}^2$. We prove that B is independent. For this, we need to show that the only solution to $a(1,1) + b(0,1) = (0,0)$ is $a = b = 0$. The coefficient matrix of the corresponding system is $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. This matrix is invertible (its determinant is 1) therefore, the system has a unique solution. Since it is a homogeneous system, this unique solution is the trivial solution. Hence, B is linearly independent, therefore it is a basis by theorem 313.*

4.5.4 Dimension of Subspaces

In the examples that follow, given the description of a subspace, we have to find its dimension. For this, we need to find a basis for it.

Example 324 *The set of 2×2 symmetric matrices is a subspace of $M_{2,2}$. Find a basis for it and deduce its dimension.*
A typical 2×2 matrix is of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. However, for this matrix to be symmetric, we must have $b = c$. Therefore, a typical 2×2 symmetric matrix is of the form $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Such a set can be spanned by $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. This can be easily seen since $\begin{bmatrix} a & b \\ b & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. For this set to be a basis, we must also prove that it is linearly independent. For this, we look at the solutions of the equation $a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. This equation is equivalent to $\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ from which it follows that the only solution is $a = b = c = 0$. Therefore, $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for the set of symmetric 2×2 symmetric matrices. Hence, the dimension of this set is 3.

Example 325 *Determine the dimension of the subspace W of $\mathbb{R}^3$ defined by $W = \{(d, c-d, c) \mid c \in \mathbb{R} \text{ and } d \in \mathbb{R}\}$.*
We notice that even though we are in $\mathbb{R}^3$, not all three coordinates of the typical

element of this subspace are independent. There is a pattern. We can write

$$\begin{aligned}(d, c-d, c) &= (0, c, c) + (d, -d, 0) \\ &= c\,(0,1,1) + d\,(1,-1,0)\end{aligned}$$

Thus, we see that the set $B = \{(0,1,1), (1,-1,0)\}$ *spans* W*. Is it a basis? We need to check if it is linearly dependent. For this, we solve the equation* $a\,(0,1,1) + b\,(1,-1,0) = (0,0,0)$*. This is equivalent to* $\left\{\begin{array}{c} b=0 \\ a-b=0 \\ a=0 \end{array}\right.$ *. We see that the only solution is* $a = b = 0$*. Therefore,* B *is independent, it follows that it is a basis for* W*. Hence,* $\dim(W) = 2$*.*

Example 326 *Determine the dimension of the subspace* W *of* $\mathbb{R}^3$ *defined by* $W = \{(x, y, 0) \,|x \in \mathbb{R}$ *and* $y \in \mathbb{R}\}$*. Give its geometric description.*
We notice that even though we are in $\mathbb{R}^3$*, not all three coordinates of the typical element of this subspace are independent. There is a pattern. We can write*

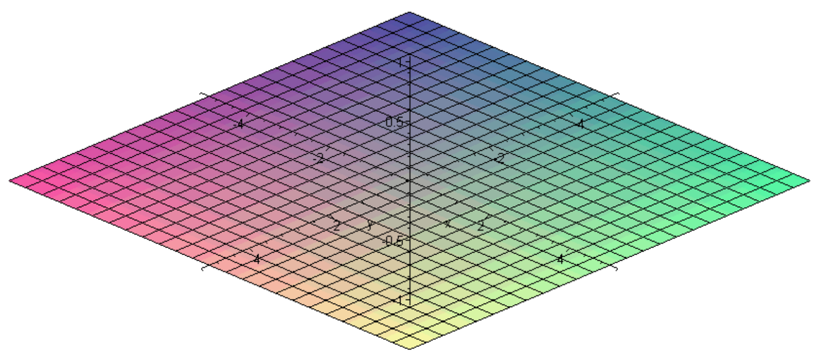
$$\begin{aligned}(x, y, 0) &= (x, 0, 0) + (0, y, 0) \\ &= x\,(1,0,0) + y\,(0,1,0)\end{aligned}$$

Thus, we see that $B = \{(1,0,0), (0,1,0)\}$ *spans* W*. It is also easy to check that it is linearly independent (left to the reader to do). Hence it is a basis for* W*. It follows that* $\dim(W) = 2$*.* W *is the set of vectors in* $\mathbb{R}^3$ *whose third coordinate is zero. These vectors only have an* x *and* y *coordinate. Therefore, the live in the x-y plane in* $\mathbb{R}^3$*. A picture of this plane is shown in figure 326.*

Next, we look at some example in which a spanning set of the subspace is given and we have to find its dimension. If the given spanning set is independent, it will be a basis and the dimension of the space will be the number of elements of the given set. Otherwise, we will have to eliminate from the spanning sets the vectors which are linear combinations of the other vectors in the given set, until we have a linearly independent set.

Example 327 *Determine the dimension of the subspace* W *of* $\mathbb{R}^4$ *spanned by* $S = \{(-1,2,5,0), (3,0,1,-2), (-5,4,9,2)\}$*.*
We must determine if S *is linearly independent. Let* $\mathbf{v}_1 = (-1,2,5,0)$*,* $\mathbf{v}_2 = (3,0,1,-2)$ *and* $\mathbf{v}_3 = (-5,4,9,2)$*. Then, it is easy to see that* $\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$*, thus proving* S *is linearly dependent. By a theorem studied before, we know we can remove* $\mathbf{v}_3$ *from* S *to obtain* $S_1 = (\mathbf{v}_1, \mathbf{v}_2)$ *and* S_1 *will span the same set as* S*. If* S_1 *is linearly independent, we are done. For this, we look at the equation* $a\mathbf{v}_1 + b\mathbf{v}_2 = \mathbf{0}$*. It is equivalent to the system*

$$\left\{\begin{array}{c} -a+3b=0 \\ 2a=0 \\ 5a+9b=0 \\ -2a+2b=0 \end{array}\right.$$


1
0.5
-0.5
-1
-4
-2
2
4
-4
-2
2
4
y
x

The only solution to this system is $a = b = 0$*. It follows that* S_1 *is independent, hence a basis for* W*. Hence,* $\dim(W) = 2$*.*

Remark 328 *Suppose we fail to see that* $\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$*, which was critical for doing this problem. What can we do? It's easy. Set up the system to see if the given set is independent. The system is* $a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3 = 0$*. This system is equivalent to*

$$\begin{cases} -a + 3b - 5c = 0 \\ 2a + 4c = 0 \\ 5a + b + 9c = 0 \\ -2b + 2c = 0 \end{cases}$$

The corresponding augmented matrix is

$$\begin{bmatrix} -1 & 3 & -5 & 0 \\ 2 & 0 & 4 & 0 \\ 5 & 1 & 9 & 0 \\ 0 & -2 & 2 & 0 \end{bmatrix}$$

Reducing it with Gaussian elimination produces

$$\begin{bmatrix} -1 & 3 & -5 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The corresponding system is

$$\begin{cases} -a + 3b - 5c = 0 \\ b - c = 0 \end{cases}$$

The solutions are

$$\begin{cases} a = -2c \\ b = c \end{cases}$$

If we write the solution in parametric form, we obtain

$$\begin{cases} a = -2t \\ b = t \\ c = t \end{cases}$$

for any real number t*. Letting* $t = 1$*, we obtain* $a = -2$*,* $b = c = 1$*. Hence, the equation* $a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3 = 0$ *becomes* $-2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = 0$ *which is equivalent to*

$$\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$$

Example 329 *We proved in an earlier section that the solution set of a homogeneous system formed a vector space. In this example, we illustrate how to find* *its dimension and a basis for it. Determine a basis and the dimension of the solution space of*

$$\begin{cases} 2x_1 + 2x_2 - x_3 + x_5 = 0 \\ -x_1 - x_2 + 2x_3 - 3x_4 + x_5 = 0 \\ x_1 + x_2 - 2x_3 - x_5 = 0 \\ x_3 + x_4 + x_5 = 0 \end{cases}$$

We begin by solving the system. Its augmented matrix is

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

Its reduced row-echelon form is

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & | & 0 \\ 0 & 0 & 1 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

Thus its solutions are

$$\begin{cases} x_1 + x_2 + x_5 = 0 \\ x_3 + x_5 = 0 \\ x_4 = 0 \end{cases}$$

If we write the solution in parametric form, we get

$$\begin{cases} x_1 = -s - t \\ x_2 = s \\ x_3 = -t \\ x_4 = 0 \\ x_5 = t \end{cases}$$

So, we have

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} &= \begin{bmatrix} -s \\ s \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -t \\ 0 \\ -t \\ 0 \\ t \end{bmatrix} \\ &= s \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

Thus, the vectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ *and* $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$. *Since they are also linearly independent (check!), they form a basis. So, the dimension of the solution space is* 2.

We finish this section with a few very important theorems we will give without proof.

Theorem 330 *Let S be a nonempty set of vectors in V.*

1. *If S is linearly independent and $\mathbf{v} \in V$ but $\mathbf{v} \notin span\,(S)$ then $S \cup \{\mathbf{v}\}$ is also linearly independent.*

2. *If $\mathbf{v} \in S$ and $\mathbf{v}$ is a linear combination of elements of S then $span\,(S) = span\,(S - \{\mathbf{v}\})$*

Corollary 331 *Let S be a finite set of vectors of a finite-dimensonal vector space V.*

1. *If S spans V but is not a basis for V, then S can be reduced to a basis for V by removing appropriate vectors from S.*

2. *If S is linearly independent but does not span V, then S can be enlarged to a basis for V by inserting appropriate vectors in S.*

Proof. *We prove each part separately.*

1. *If S is not a basis, then it is not independent. It means some vectors of S are linear combination of other vectors in S. By part 2 of the theorem, such vectors can be removed. This way, we can remove all the vectors of S which are linear combination of other vectors. When none such vectors are left, S will be linearly independent hence a basis.*

2. *Here, we use part 1 of the theorem. If $span\,(S) \neq V$, we can pick a vector in V not in $span\,(S)$ and add it to S. S will still be linearly independent. We continue to add vectors until $span\,(S) = V$.*

■

Theorem 332 *If W is any subspace of a finite-dimensional vector space V then* $\dim(W) \leq \dim(V)$. *Moreover, if* $\dim(W) = \dim(V)$ *then $W = V$.*

4.5.5 Summary

- Know and understand the definition of a basis for a vector space.
- Know what the dimension of a vector space is.
- Know what the coordinates of a vector relative to a given basis are.
- Given a set of vectors in a vector space, be able to tell if that set is a basis for the vector space.
- Know the standard basis for common vector spaces such as $\mathbb{R}^n$, M_{nn}, P_n for every positive integer n.
- Be able to find the basis of subspaces given the description of a subspace.
- Be able to find the coordinates of any vector relative to a given basis.
- Be able to find the dimension of a vector space.
- Know and understand the difference between a finite-dimensional and infinite dimensional vector space.

4.5.6 Problems

Exercise 333 *Let V be a vector space, and S a subset of V containing n vectors. What can be said about* $\dim(V)$ *if we know that S spans V?*

Exercise 334 *Let V be a vector space, and S a subset of V containing n vectors. What can be said about* $\dim(V)$ *if we know that S is linearly independent?*

Exercise 335 *Let V be a vector space of dimension n. Can a subset S of V containing less than n elements span V?*

Exercise 336 *Let V be a vector space of dimension n. Can a subset S of V containing less than n elements be dependent? If yes, is it always dependent?*

Exercise 337 *Same question for independent.*

Exercise 338 *Let V be a vector space, and S a subset of V containing n vectors. If S is linearly independent, will any subset of S be linearly independent? why?*

Exercise 339 *Let V be a vector space, and S a subset of V containing n vectors. If S is linearly dependent, will any subset of S be linearly dependent? why?*

Exercise 340 *What is the dimension of $C(-\infty,\infty)$ and why?*

Exercise 341 *Do # 1, 2, 3, 4, 6, 7, 8, 10, 12, 19, 21, 23, 24, 30, 32 on pages 263 - 265*

DM559

Linear and Integer Programming

Lecture 8
Change of Basis

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

Resume

- Linear dependence and independence
- Determine linear dependency of a set of vectors, ie, find non-trivial lin. combination that equal zero
- Basis
- Find a basis for a linear space
- Dimension (finite, infinite)

Outline

Coordinates

Recall:

Definition (Coordinates)

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis of a vector space V, then

- any vector $\mathbf{v} \in V$ can be expressed uniquely as $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$
- and the real numbers $\alpha_1, \alpha_2, \dots, \alpha_n$ are the coordinates of $\mathbf{v}$ wrt the basis S.

To denote the coordinate vector of $\mathbf{v}$ in the basis S we use the notation

$$[\mathbf{v}]_S = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}_S$$

- In the standard basis the coordinates of $\mathbf{v}$ are precisely the components of the vector $\mathbf{v}$: $\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n$
- How to find coordinates of a vector $\mathbf{v}$ wrt another basis?

Transition from Standard to Basis B

Definition (Transition Matrix)

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbb{R}^n$. The coordinates of a vector $\mathbf{x}$ wrt B, $\mathbf{a} = [a_1, a_2, \ldots, a_n]^T = [\mathbf{x}]_B$, are found by solving the linear system:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n = \mathbf{x} \qquad \text{that is} \qquad [\mathbf{v}_1\ \mathbf{v}_2\ \cdots \mathbf{v}_n][\mathbf{x}]_B = \mathbf{x}$$

We call P the matrix whose columns are the basis vectors:

$$P = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots \mathbf{v}_n]$$

Then for any vector $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x} = P[\mathbf{x}]_B$$ transition matrix from B coords to standard coords

moreover P is invertible (columns are a basis):

$$[\mathbf{x}]_B = P^{-1}\mathbf{x}$$ transition matrix from standard coords to B coords

Example

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\} \qquad [\mathbf{v}]_B = \begin{bmatrix} 4 \\ 1 \\ -5 \end{bmatrix}$$

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

$\det(P) = 4 \neq 0$ so B is a basis of $\mathbb{R}^3$
We derive the standard coordinates of $\mathbf{v}$:

$$\mathbf{v} = 4 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} - 5 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -9 \\ -3 \\ -5 \end{bmatrix}$$

$$\mathbf{v} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 2 \\ -1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ -5 \end{bmatrix}_B = \begin{bmatrix} -9 \\ -3 \\ -5 \end{bmatrix}$$

Example (cntd)

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}, \qquad [\mathbf{x}]_S = \begin{bmatrix} 5 \\ 7 \\ -3 \end{bmatrix}$$

We derive the B coordinates of vector $\mathbf{x}$:

$$\begin{bmatrix} 5 \\ 7 \\ -3 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + a_2 \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} + a_3 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

either we solve $P\mathbf{a} = \mathbf{x}$ in $\mathbf{a}$ by Gaussian elimination or we find the inverse P^{-1}:

$$[\mathbf{x}]_B = P^{-1}\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}_B \qquad \text{check the calculation}$$

What are the B coordinates of the basis vector? ($[1, 0, 0], [0, 1, 0], [0, 0, 1]$)

Change of Basis

Since $T(\mathbf{x}) = P\mathbf{x}$ then $T(\mathbf{e}_i) = \mathbf{v}_i$, ie, T maps standard basis vector to new basis vectors

Example

Rotate basis in $\mathbb{R}^2$ by $\pi/4$ anticlockwise, find coordinates of a vector wrt the new basis.

$$A_T = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Since the matrix A_T rotates $\{\mathbf{e}_1, \mathbf{e}_2\}$, then $A_T = P$ and its columns tell us the coordinates of the new basis and $\mathbf{v} = P[\mathbf{v}]_B$ and $[\mathbf{v}]_B = P^{-1}\mathbf{v}$. The inverse is a rotation clockwise:

$$P^{-1} = \begin{bmatrix} \cos(-\frac{\pi}{4}) & -\sin(-\frac{\pi}{4}) \\ \sin(-\frac{\pi}{4}) & \cos(-\frac{\pi}{4}) \end{bmatrix} = \begin{bmatrix} \cos(\frac{\pi}{4}) & \sin(\frac{\pi}{4}) \\ -\sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Example (cntd)

Find the new coordinates of a vector $\mathbf{x} = [1, 1]^T$

$$[\mathbf{x}]_B = P^{-1}\mathbf{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}$$

Change of basis from B to B'

Given an old basis B of $\mathbb{R}^n$ with transition matrix P_B,
and a new basis B' with transition matrix $P_{B'}$,
how do we change from coords in the basis B to coords in the basis B'?

coordinates in $B \xrightarrow{\mathbf{v}=P_B[\mathbf{v}]_B}$ standard coordinates $\xrightarrow{[\mathbf{v}]_{B'}=P_{B'}^{-1}\mathbf{v}}$ coordinates in B'

$$[\mathbf{v}]_{B'} = P_{B'}^{-1} P_B [\mathbf{v}]_B$$

$$M = P_{B'}^{-1} P_B = P_{B'}^{-1} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] = [P_{B'}^{-1}\mathbf{v}_1 \quad P_{B'}^{-1}\mathbf{v}_2 \quad \dots \quad P_{B'}^{-1}\mathbf{v}_n]$$

i.e., the columns of the transition matrix M from the old basis B to the new basis B' are the coordinate vectors of the old basis B with respect to the new basis B'

Change of basis from B to B'

Theorem

If B and B' are two bases of $\mathbb{R}^n$, with

$$B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

then the transition matrix from B coordinates to B' coordinates is given by

$$M = \begin{bmatrix} [\mathbf{v}_1]_{B'} & [\mathbf{v}_2]_{B'} & \cdots & [\mathbf{v}_n]_{B'} \end{bmatrix}$$

(i.e., the columns of the transition matrix M from the old basis B to the new basis B' are the coordinate vectors of the old basis B with respect to the new basis B')

Example

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} \qquad B' = \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix} \right\}$$

are basis of $\mathbb{R}^2$, indeed the corresponding transition matrices from standard basis:

$$P = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \qquad Q = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$$

have $\det(P) = 3$, $\det(Q) = 1$. Hence, lin. indep. vectors.
We are given

$$[\mathbf{x}]_B = \begin{bmatrix} 4 \\ -1 \end{bmatrix}_B$$

find its coordinates in B'.

Example (cntd)

1. find first the standard coordinates of $\mathbf{x}$

$$\mathbf{x} = 4\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}-1\\1\end{bmatrix} = \begin{bmatrix}1 & -1\\2 & 1\end{bmatrix}\begin{bmatrix}4\\-1\end{bmatrix} = \begin{bmatrix}5\\7\end{bmatrix}$$

   and then find B' coordinates:

$$[\mathbf{x}]_{B'} = Q^{-1}\mathbf{x} = \begin{bmatrix}2 & -5\\-1 & 3\end{bmatrix}\begin{bmatrix}5\\7\end{bmatrix} = \begin{bmatrix}-25\\16\end{bmatrix}_{B'}$$

2. use transition matrix M from B to B' coordinates:
   $\mathbf{v} = P[\mathbf{v}]_B$ and $\mathbf{v} = Q[\mathbf{v}]_{B'}$ $\rightsquigarrow$ $[\mathbf{v}]_{B'} = Q^{-1}P[\mathbf{v}]_B$:

$$M = Q^{-1}P = \begin{bmatrix}2 & -5\\-1 & 3\end{bmatrix}\begin{bmatrix}1 & -1\\2 & 1\end{bmatrix} = \begin{bmatrix}-8 & -7\\5 & 4\end{bmatrix}$$

$$[\mathbf{x}]_{B'} = \begin{bmatrix}-8 & -7\\5 & 4\end{bmatrix}\begin{bmatrix}4\\-1\end{bmatrix} = \begin{bmatrix}-25\\16\end{bmatrix}_{B'}$$

Practice Problems S5 (Diagonalization)

1. Let A be an $n \times n$ matrix and $0 \neq k \in \mathbb{R}$. Prove that a number λ is an eigenvalue of A iff $k\lambda$ is an eigenvalue of kA.

2. Prove that if λ is an eigenvalue of a square matrix A, then λ^5 is an eigenvalue of A^5.

3. By inspection, find the eigenvalues of

$$\text{(a)} \quad A = \begin{bmatrix} 3 & 1 & 4 \\ 0 & -2 & 2 \\ 0 & 0 & 5 \end{bmatrix}; \quad \text{(b)} \quad B = \begin{bmatrix} -2 & 0 & 0 \\ 1 & 2 & 5 \\ 4 & 0 & 4 \end{bmatrix}$$

4. Compute $P^{-1}AP$ and then A^n if $A = \begin{bmatrix} 6 & -5 \\ 2 & -1 \end{bmatrix}$ and $P = \begin{bmatrix} 5 & 1 \\ 2 & 1 \end{bmatrix}$

5. (Diagonalization) Find the characteristic polynomial, eigenvalues and an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix if $A = \begin{bmatrix} 3 & 1 & 1 \\ -4 & -2 & -5 \\ 2 & 2 & 5 \end{bmatrix}$.

6. Determine whether the following matrices are diagonalizable or not: (a) $A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$; (b) $B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$; (c) $C = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$.

Solutions

1. It follows from $\det(k\lambda I_3 - k\,A) = \det\big(k(\lambda I_3 - A)\big) = k^n \det(\lambda I_3 - A)$.

2. Assume that λ is an eigenvalue of A, i.e., $AX = \lambda X$ for some nonzero vector X. It follows $A^5X = A^4(AX) = A^4(\lambda X) = \lambda A^4 X = \cdots = \lambda^5 X$. Therefore, λ^5 is an eigenvalue of A^5.

3. (a) The main diagonal entries of any triangular matrix are the eigenvalues.

   (b) A main diagonal entry in row or column with only zeros except possibly the main diagonal entry itself, is an eigenvalue.

4. $P^{-1} = \begin{bmatrix} 5 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \dfrac{1}{3}\begin{bmatrix} 1 & -1 \\ -2 & 5 \end{bmatrix}$. So, $P^{-1}AP = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} =$ $\mathrm{diag}(4,1)$. It follows that $A = P\mathrm{diag}(4,1)P^{-1}$. Therefore,

$$\begin{aligned} A^n &= P\mathrm{diag}(4^n,1)P^{-1} \\ &= \frac{1}{3}\begin{bmatrix} 5 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 4^n & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -2 & 5 \end{bmatrix} \\ &= \begin{bmatrix} 5/3\,4^n - 2/3 & -5/3\,4^n + 5/3 \\ 2/3\,4^n - 2/3 & -2/3\,4^n + 5/3 \end{bmatrix}. \end{aligned}$$

5. The characteristic polynomial of A is

$$\det(xI_3 - A) = \begin{vmatrix} x-3 & -1 & -1 \\ 4 & x+2 & 5 \\ -2 & -2 & x-5 \end{vmatrix} = (x-1)(x-2)(x-3).$$

   To simplify the computation of this determinant, subtract column 2 from column 1 and factor x-2 out of column 1 from the new determinant, then add row 1 to row 2; finally subtract column 3 from column 2. Thus A has three eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = 3$. Eigenvectors: The homogeneous systems $(I_3 - A)X = 0$, $(2I_3 - A)X = 0$ and $(3I_3 - A)X = 0$ have basic solutions $X_1 = [1\ -3\ 1]^T$, $X_2 = [1\ -1\ 0]^T$ and $X_3 = [0 - 1\,1]^T$, respectively. There are basic eigenvectors correponding to the respective eigenvalues. The matrix $P = [X_1\,X_2\,X_3] =$ $\begin{bmatrix} 1 & 1 & 0 \\ -3 & -1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$ diagonalizes A with $P^{-1}AP = \mathrm{diag}(1,2,3)$.

6. (a) Since the 2×2 matrix A has two distinct (or simple, i.e., with multiplicity 1) eigenvalues $\lambda_1 = -5$ and $\lambda_2 = 2$, A is diagonalizable.

   (b) The matrix B has eigenvalue $\lambda = -1$ with **multiplicity** 2. The matrix B is diagonalizable if there are two basic eigenvectors corresponding $\lambda = -1$. There are basic solutions to the homogeneous system $(-I_3 - B)X = 0$. So, $X_1 = [-1\ 1, 0]$ and $X_2 = [-1, 0, 1]$ are **two basic eigenvectors** corresponding to $\lambda = -1$. The matrix B is diagolizable.

   (c) The matrix C has an eigenvalue $\lambda = 1$ of multiplicity 2. For C to be diagonalizable, there must be two basic eigenvectors corresponding to $\lambda = 1$. But the homogeneous system $(I_3 - C)X = 0$ has only one basic solution $X = [0\ 1\ 0]^T$. Therefore, C is not diagonalizable.

Exercise sheet n° 7

Eigenvalues and eigenvectors

1. Compute the eigenvalues and the eigenvectors of the following matrices.

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & -4 \\ 4 & 1 & -4 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad C = \begin{pmatrix} 4 & 1 & 2 \\ -1 & 1 & -1 \\ -2 & -1 & 0 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 1 & -2 \\ 0 & 2 & -2 \\ 1 & 0 & 0 \end{pmatrix}.$$

Which of the previous matrices is diagonalizable ?

2. Let

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

(i) Show that A is diagonalizable and find a matrix P diagonalizing A. Compute A^n for all $n \geq 1$.

(ii) Consider the sequences (u_n), (v_n) and (w_n) defined by the initial values $u_0 = v_0 = 1$, $w_0 = 2$ and the following recursive relations :

$$u_{n+1} = 3u_n - v_n + w_n \qquad v_{n+1} = 2v_n \qquad w_{n+1} = u_n - v_n + 3w_n.$$

Compute u_n, v_n and w_n.

3. Let $a, b \in \mathbb{R}$ and

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

Compute A^n for all $n \geq 1$.

4. Let $A \in \mathcal{M}_n(\mathbf{C})$ be a nilpotent matrix, *i.e.* there exists $p \in \mathbf{N}^*$ such that $A^p = 0$. Show that the only eigenvalue of A is 0. Does the converse hold ?

5. Let $A \in \mathcal{M}_n(\mathbf{C})$. Show that the determinant of A is equal to the product of its eigenvalues (counted with multiplicity) and that the trace of A is equal to the sum of its eigenvalues (also counted with multiplicity).

6. Let $A, B \in \mathcal{M}_n(\mathbf{C})$. Show that AB and BA have the same set of eigenvalues, each of them with the same multiplicity.

7. The Italian mathematician Leonardo Fibonacci (c. 1175 – c. 1250) was the first to study the sequence of integers given by 1, 1, 2, 3, 5, 8,. . . and that has his name. It is recursively defined by

$$f_0 = 1,\ f_1 = 1,\ f_{k+1} = f_k + f_{k-1},\ k = 1, 2, ...$$

1. Let $x^{(k)} = (f_{k+1}; f_k)$. Write this relations in matricial form
$$x^{(k+1)} = Ax^{(k)},\ k = 0, 1, ... ,\ \ x^{(0)} = (1; 1),$$
, where A is a matrix to be determiend .
2. Compute the eigenvalues and the eigenvectors of A. Is the matrix LA diagonalizable ?
3. Find an explicit formula for the k-th element of the Fibonacci sequence.

8. Compute the eigenvalues of the matrix

$$A = \begin{pmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 4 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Consider now the system of differential equations given by

$$\dot{x}(t) = x(t) + 4z(t), \qquad \dot{y}(t) = y(t) + 4w(t),$$
$$\dot{z}(t) = x(t) + z(t), \qquad \dot{w}(t) = y(t) + w(t).$$

Compute the general solution of the system, and then the particular solution satisfying the conditions $x(0) = y(0) = z(0) = 0, w(0) = 2$.

9. Let $a \in I\!R$. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & a \\ 0 & -1 & 2 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}.$$

(i) Compute the characteristic polynomial of A and the corresponding eigenvalues.

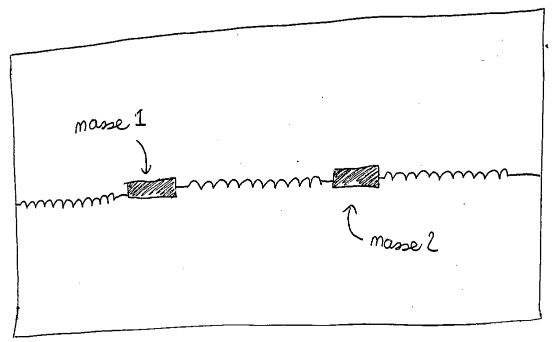
(ii) Compute the eigenspaces of A associated to each eigenvalue. Determine all the values of the parameter a such that the matrix A is diagonalizable. And triangularizable ?

(iii) Assume that $a = 0$ and consider the system of differential equations given by

$$\dot{x}(t) = x(t) + y(t) - z(t), \qquad \dot{y}(t) = y(t),$$
$$\dot{z}(t) = -y(t) + 2z(t), \qquad \dot{w}(t) = x(t) + z(t) + 2w(t).$$

Compute the general solution of the system, and then the particular solution satisfying the conditions $x(0) = y(0) = w(0) = 0, z(0) = 1$.

10. Consider the system formed by two masses and three springs with the same force constant $k > 0$ as indicated in the following diagram :



Assume that the two are longitudinally separated from the equilibrium position (using an external force) at time $t = 0$ and the systems stars moving freely. Let $x_1(t)$ (resp. $x_2(t)$) the (signed) longitudinal distance of the first (resp. second) mass from its equilibrium position at t. We assume that the two springs obey Hooke's law, so

$$\begin{aligned} mx_1'' &= -2kx_1 + kx_2, \\ mx_2'' &= kx_1 - 2kx_2. \end{aligned}$$

1. Rewrite these equations in matrix form :

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}'' = M \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

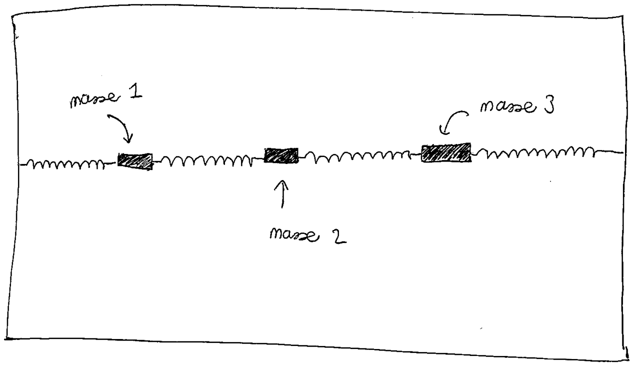
   where M s a matrix to be determined.
2. Diagonalize M and compute the general solution of this equation.
3. What are normal modes of oscillation of the system ?

We consider now the analogous situation with 3 masses and 4 springs.

1. Write down the equations of motion of this system.
2. Rewrite them in matrix form :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}'' = M \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

   where M s a matrix to be determined.
3. Diagonalize M and compute the general solution of this equation.
4. What is the normal mode of oscillation of the system corresponding to the eigenvalue $2k/m$.



Eigenvalues of symmetric matrices and singular value decomposition

11. Show that the following matrices are orthogonal and compute their eigenvalues.

$$(a) \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (b)\ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad (c) \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

12. Show that the following matrices are symmetric and diagonalize them by means of an orhtogonal matrix.

$$(a) \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \quad (b) \begin{bmatrix} 2 & 36 \\ 36 & 23 \end{bmatrix} \quad (c) \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

13. Compute the singular value decomposition of the following matrices.

$$(a) \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad (b) \begin{bmatrix} -2 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \quad (c) \begin{bmatrix} 0 & -2 & 0 \\ 2 & 0 & 0 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 2 \end{bmatrix}$$

14. Let $A = U\Sigma V^t$ be the singular value decomposition of a matrix $A \in I\!R^{m\times n}$ of rank $p < \min(m,n)$. We denote by $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p$ the nonzero singular values of A, and by $\bar{U} = [u_1, ..., u_p]$ and $\bar{V} = [v_1, ..., v_p]$ the left and right associated orthogonal matrices, respectively. We have thus

$$A = \bar{U}\,\mathrm{diag}(\sigma_1, ..., \sigma_p)\,\bar{V}^t$$

(minimal decomposition).

Let A^+ be the matrix of size $n \times m$ given by

$$A^+ = \bar{V}\,\mathrm{diag}(\sigma_1^{-1}, ..., \sigma_p^{-1})\,\bar{U}^t.$$

It is called the *pseudo-inverse matrix of* A.

1. What is the size of A^+? Express A^+A and AA^+ in terms of the singular value decomposition of A. Verify that $A\,A^+A = A$ and $A^+A\,A^t = A^t\,AA^+ = A^t$. Explain their meaning.
2. Let A be a matrix of size $m \times n$ and rank n, and let b be a vector column of size m. Assume that the system $Ax = b$ has a solution x. Show that the solution $x \in I\!R^n$ satisfies $x = A^+b$.

Outline

- Mathematical background
- PCA
- SVD
- Some PCA and SVD applications
- Case study: LSI

Mathematical Background

Variance

If we have **one** dimension:

- English: The average square of the distance from the mean of the data set to its points
- Definition: *Var(X)*= $E(X - E(X))^2 \; = \; E(x^2) - (E(X))^2$
- Empirical: *Var(x)*= $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$

Many datasets have more than one dimension. Example: we might have our data set both the height of all students and the mark they received, and we want to see if the height has an effect on the mark.

Mathematical Background

Covariance

Always measured between **two** dimensions.

- English: For each data item, multiply the difference between the x value and the mean of x, by the difference between the y value and the mean of y.
- Definition: *cov(X,Y)= E[(X-E(X))(Y-E(Y))] = E(X.Y)-E(X).E(Y)*
- Empirical: $cov\ (X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$

  the inner product of values on the X dimension with the values on the Y dimension (after subtracting the means): $\frac{1}{N}\ (X^T.Y)$

Mathematical Background

Covariance properties

- $cov(X,X)=Var(X)$
- $cov(X,Y)=cov(Y,X)$
- If X and Y are independent (uncorrelated) → $cov(X,Y)=0$
- If X and Y are correlated (both dimensions increase together) → $cov(X,Y)>0$
- If X and Y are anti-correlated (one dimension increases, the other decreases) → $cov(X,Y)<0$

Correlation

Is a scaled version of covariance: $cor\,(X,Y) = \dfrac{cov\,(X,Y)}{\sigma_X\,\sigma_Y}$

Always $-1 <= cor(X,Y) <= 1$

Mathematical Background

Covariance Matrix

Recall that covariance is a measure between two dimensions.

For example, if we have 3 dimensional data set (dimensions x, y, z), we should calculate *cov(x,y)*, *cov(y,z)*, and *cov(x,z)*.

For n dimensional data, we calculate n!/(n-2)!*2 = n(n-1)/2 different covariance values.

The covariance matrix for a set on data with n dimensions is:

C(n x n)= (c[i,j]=c[j,i]=cov(Dim[i],Dim[j])

The covariance matrix for a 3 dimensional data is

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

Mathematical Background

Covariance Matrix Properties

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

- C is square and symmetric matrix.
- The diagonal values are the variance for each dimension and the off-diagonal are the covariance between measurement types.
- Large term in the diagonal correspond to interesting dimensions, whereas large values in the off-diagonal correspond to high correlations (redundancy).

Because we want to minimize the correlation (redundancy) and maximize the variance, we would like to have a diagonal covariance matrix.

Mathematical Background

eigenvectors

Example: $$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

if we thought of the squared matrix as a transformation matrix, then multiply it with the Eigenvector don't change its direction.

eigenvalues and eigenvectors always come in pairs. In the example: 4 is the eigenvalue of our eigenvector.

No matter what multiple of the eigenvector we took, we get the same eigenvalue.

Example: $$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

Mathematical Background

eigenvectors

The length of a vector doesn't affect whether it's an eigenvector or not, whereas the direction does. So to keep eigenvectors standard, we scale them to have length 1.

So we scale our vector: $$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \div \sqrt{13} = \begin{pmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \end{pmatrix}$$

eigenvectors properties

1. eigenvectors can only be found for **square** matrices.

2. Not every square matrix has eigenvectors.

Mathematical Background

eigenvectors properties (continue)

3. If we scale the eigenvector by an amount, we will still get the same eigenvalue *(as we saw)*.

4. **Symmetric** matrices S(n x n) also satisfies two properties:

    I. has exactly n eigenvectors.

    II. All the eigenvectors are orthogonal (perpendicular). This is important because we can express the data in term of eigenvectors, instead of expressing them in the original space.

We say that eigenvectors are *orthonormal*, which means orthogonal and has length 1.

What are the eigenvectors of the identity matrix?

Any vector is an eigenvector to the identity matrix.

Outline

- Mathematical background
- PCA
- SVD
- Some PCA and SVD applications
- Case study: LSI

PCA and SVD

PCA: Principle Components Analysis, also known as KLT (Karhunen-Loeve Transform).

SVD: Singular Value Decomposition.

SVD and PCA are closely related.

Why we use SVD and PCA?

- A powerful tool for analyzing data and finding patterns.
- Used for compression. So you can reduce the number of dimensions without much loss of information.

PCA

Objective: project the data onto a lower dimensional linear space such that the variance of the projected data is maximized. Equivalently, it is the linear projection that minimizes the average projection cost (mean squared distance between the data points and their projections).

Different from the feature subset selection !!!

Problem to solve: In high dimensional space, we need to learn a large number of parameters. Thus if the dataset is small, this will result in large variance and over-fitting.

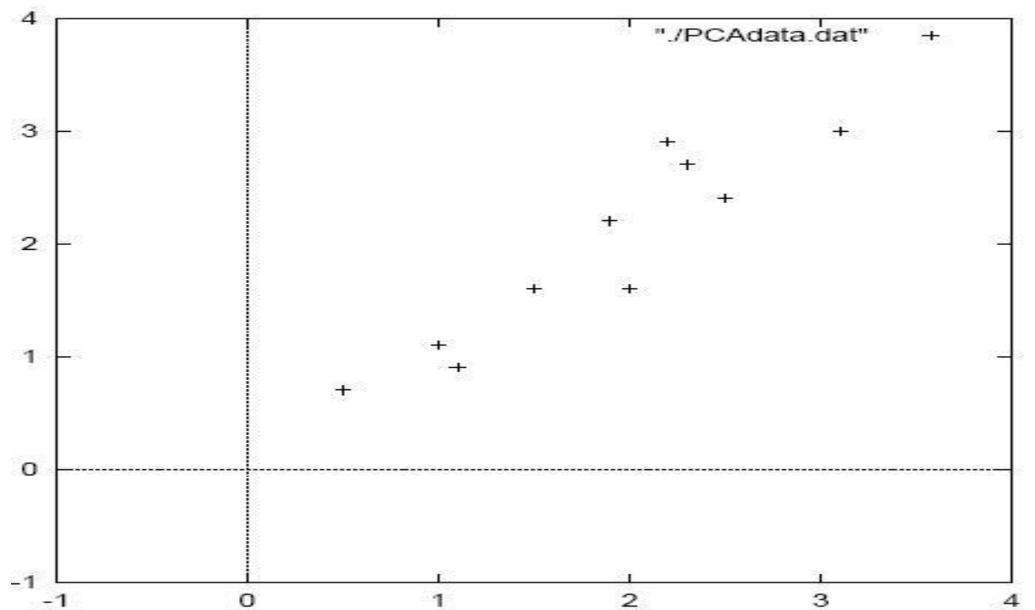
We want to represent the vector *x* in a different space (p dimensional) using a set of orthonormal vectors *U (*where u_i is a principle component).

PCA

Method to perform PCA on a data

- <u>Step 1: get some data</u>

  Let A (N,n) be the data matrix: N is the number of data points, n is the number of dimensions. It could represent N patients with n numerical symptoms each (blood pressure, cholesterol level etc) or N documents with n terms in each document (used in IR).



PCA

Method to perform PCA on a data

- Step 2: Subtract the mean

  Intuitively, we translate the origin to the center of gravity. We obtain the matrix B (N,n):

  $B = [b_{i,j}] = [a_{i,j} - \overline{a_{*,j}}].$

  This produces a zero mean data (the column averages are zero).

- Step 3: Calculate the covariance matrix C.

  $C(n,n)=\frac{1}{N}[B^T(n,N) \text{ x } B(N,n)]$: T is transpose.

  Since our data is 2D, the covariance matrix will be (2 x 2).

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

Notice that the non-diagonal elements are positive, why?

PCA

Method to perform PCA on a data

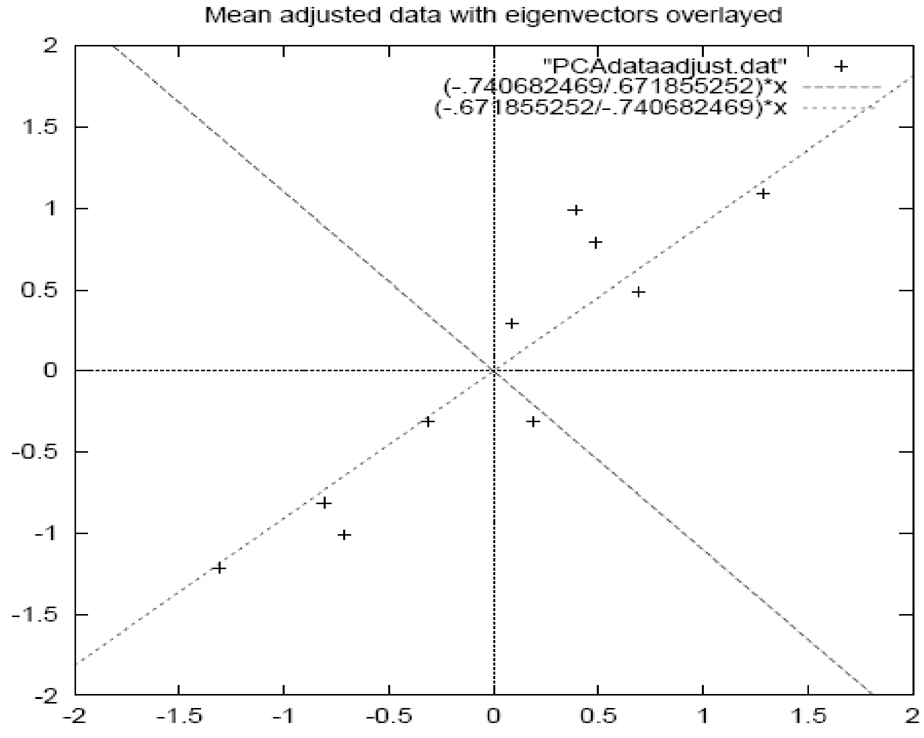
- Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix

Since the covariance matrix is square, we can calculate the eigenvectors

$\lambda_1 \approx 1.28$, $V_1 \approx [-0.677\ \ -0.735]^T$, $\lambda_2 \approx 0.49$, $V_2 \approx [-0.735\ \ 0.677]^T$

Notice that V_1 and V_2
Are orthonormal, why?

$V_1 . V_2 = 0$
$|V_1|=1$
$|V_2|=1$



PCA

Method to perform PCA on a data

- Step 5: Choosing components and forming a feature vector

  Objective: project the *n* dimensional data on a *p* dimensional sub-space ($p \leq n$), minimizing the error of the projections (sum of squared difference).

  Here is where we reduce the dimensionality of the data (for example, to do compression).

  How: Order the eigenvalues from highest to lowest to get the components in order of significance. Project on the *p* eigenvectors that corresponds to the highest *p* eigenvalues.

PCA

Method to perform PCA on a data

- Step 5 (continue)

The eigenvector with the highest eigenvalue is the *principle component* of the data.

if we are allowed to pick only one dimension to project the data on it, then the principle component is the best direction.

the PC of our example is :

$$\begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix}$$

PCA

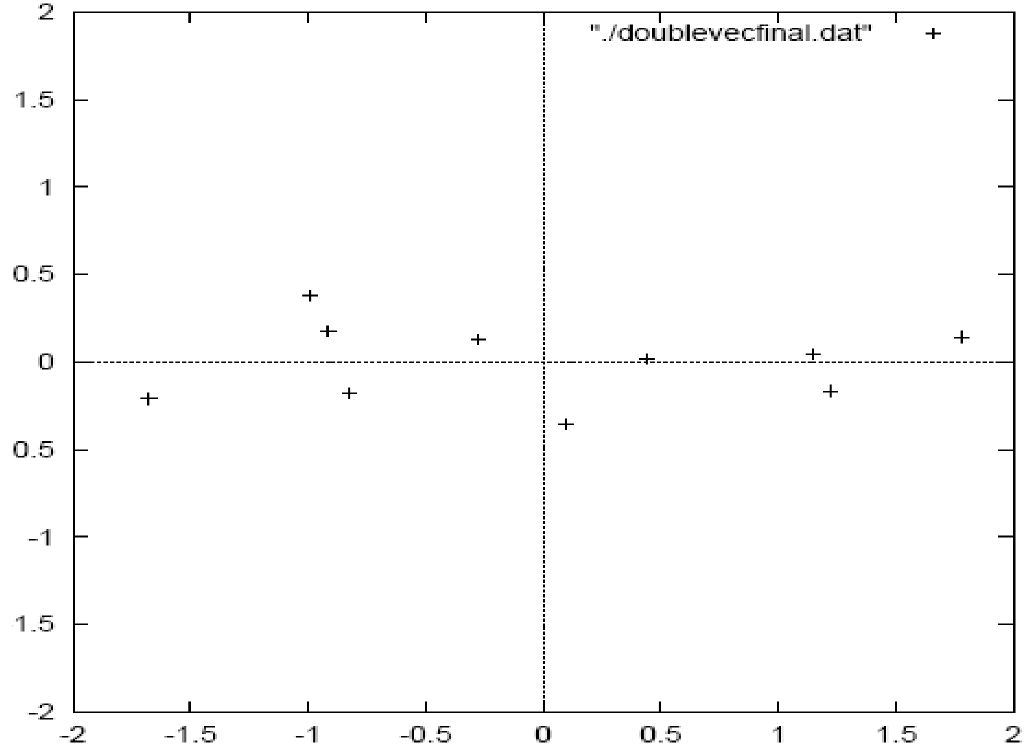
Method to perform PCA on a data

➢ <u>Step 6: Derive the new data set</u>

Let's denote Feature space matrix by U(n x p): where the columns are the eigenvectors. Let the final data be F.

F(N,p)=B(N,n) x U(n,p)

The final data F resides in a p-dimensional feature space.

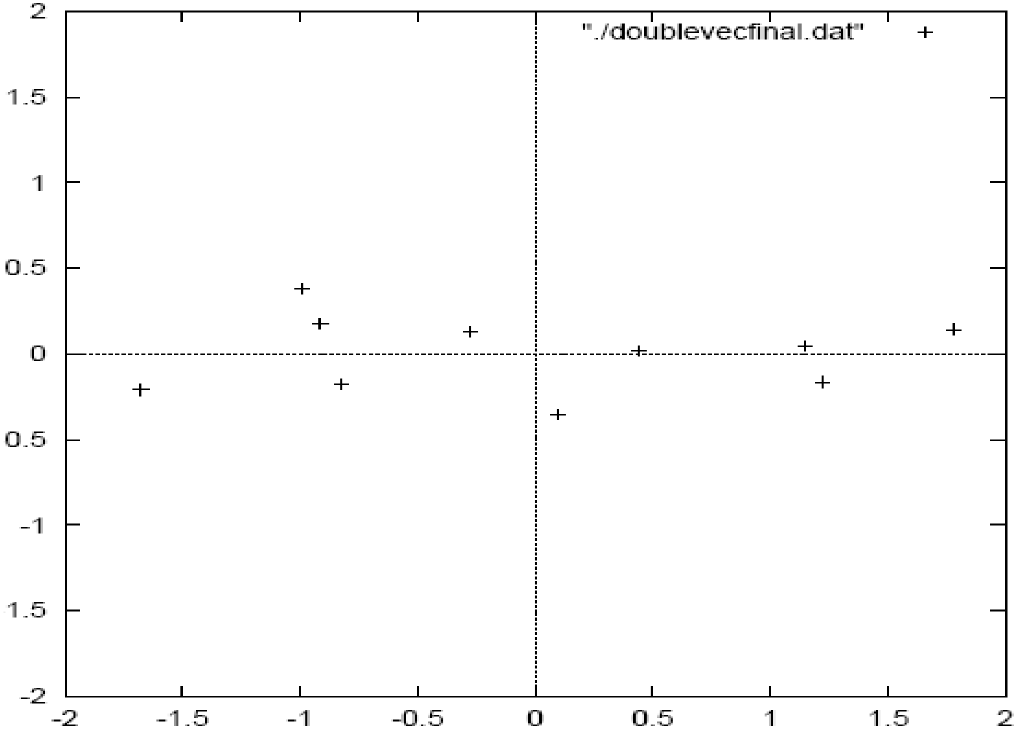


PCA

Method to perform PCA on a data

- <u>Step 6: (continue)</u>

  For our example, if we keep both eigenvectors, we get the original data, rotated so that The eigenvectors are the axes. (we've lost no information)



PCA

Method to perform PCA on a data

- Step 7 (optional): getting the old data back. (if we are doing compression.)
  If we keep all eigenvectors, we will get exactly the same data.

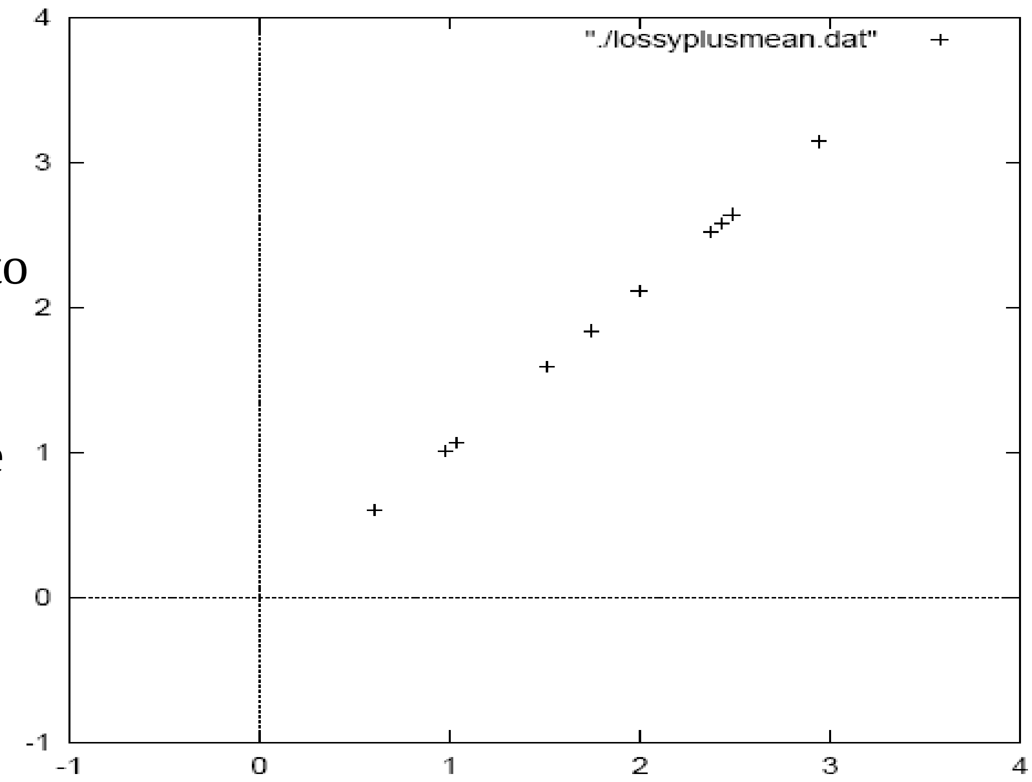
We know F=B x U

→B = F x U^{-1}

→B = F x U^T : (*proof later)*

To get A, we add the mean to vector to B.

Notice: The variance along the Other Component has gone (a lossy compression)



PCA and FLD

FLD: Fisher's Linear Discriminant.

Is a supervised learning method (utilizes the class label) that is used to select the projection that maximizes the class separation.

Specifically: we have $y=w^T x$, FLD tries to adjust the components of w such that it maximizes the distance between the projected means, while minimizing the variance within each class.

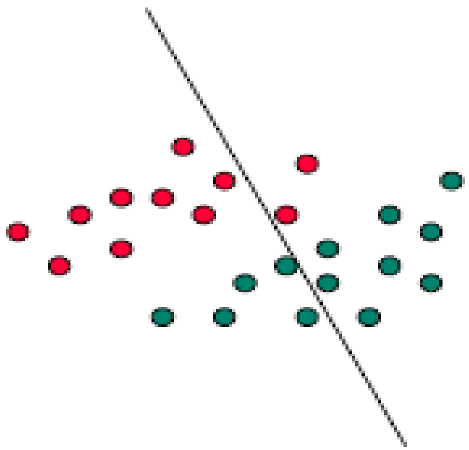
$$J(W) = \frac{m_2 - m_1}{s_1^2 + s_2^2}$$ where m_1 and m_2 are the projected means and

$$s_k^2 = \sum_{i \in C_k} (y_i - m_k)^2$$

PCA and FLD

PCA is unsupervised learning technique that finds the dimensions useful to represent the data, but maybe bad for discrimination between different classes.

Example: Project 2D data into 1D. We have two classes: red and green.



What direction will PCA choose?

Outline

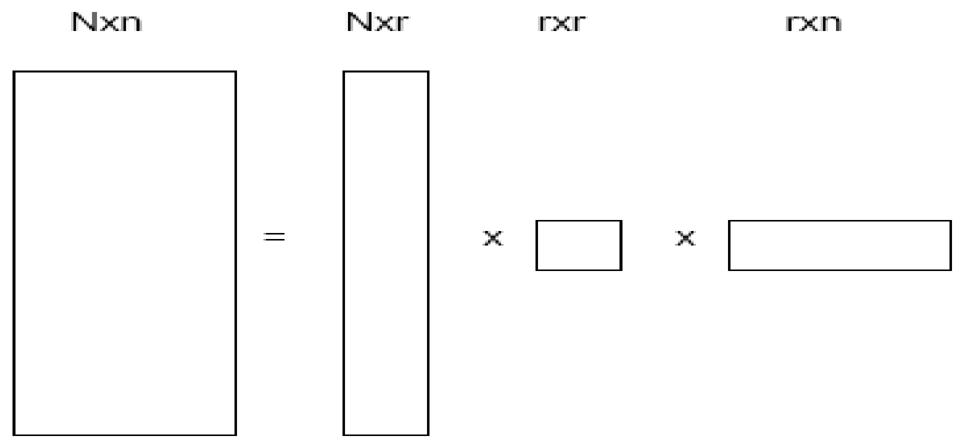
- Mathematical background
- PCA
- SVD
- Some PCA and SVD applications
- Case study: LSI

SVD

The eigenvalues and eigenvectors are defined for squared matrices. For rectangular matrices, a closely related concept is Singular Value Decomposition (SVD).

Theorem: Given an N x n real matrix A, we can express it as:

$$\mathbf{A} = \mathbf{U} \times \Lambda \times \mathbf{V}^{\mathbf{T}}$$



where U is a column-orthonormal N x r matrix, r is the rank of the matrix A (number of linearly independent rows or columns), Λ is a diagonal r x r matrix where the elements are sorted in descending order, and V is a column-orthonormal n x r matrix.

SVD decomposition for a matrix is unique.

SVD

The values of the diagonal Λ are called singular values. (we will see later that they correspond to the square root of the eigenvalues of the covariance matrix).

Theorem: the inverse of an orthonormal matrix is its transpose.

Proof: we know that $(A^T A)_{ij} = a_i{}^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Therefore A^T x $A = I$ where I is the identity matrix.

From the definition of A^{-1} : A^{-1} x $A = I$

$\rightarrow A^{-1} = A^T$

U^T x $U = I$ and V^T x $V = I$

A also can be written using spectral decomposition as:

$$A = \lambda_1 U_1 V_1^T + \lambda_2 U_2 V_2^T + \ldots + \lambda_r U_r V_r^T$$

SVD

Theorem: if S is a real and symmetric ($S=S^T$) matrix then

$\mathbf{S} = \mathbf{U} \times \mathbf{\Lambda} \times \mathbf{U}^{\mathbf{T}}$. *Where the columns of U are the eigenvectors, and Λ is a diagonal matrix with values corresponding to eigenvalues.*

Proof: let U be the matrix of eigenvectors placed in the columns:

$U=[u_1\ u_2\ \ldots\ u_n]$

We can write: $S \times U=U \,\mathbf{x}\, \mathbf{\Lambda}$

$[S \times u_1\ S \times u_2\ \ldots\ S \times u_n]=[\lambda_1.u_1\ \lambda_2.u_2\ \ldots\ \lambda_n.u_n]$ which is the definition of the eigenvectors.

Therefore: $S= U \times \Lambda \times U^{-1}$

Because U is orthonormal → $U^{-1} = U^T$

→$\mathbf{S} = \mathbf{U} \times \mathbf{\Lambda} \times \mathbf{U}^{\mathbf{T}}$

SVD

$$\begin{array}{cc} & \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{array}{c} \\ \text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \underbrace{\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}}_{\text{doc-to-concept similarity matrix}} \times \underbrace{\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}}_{\text{concepts strengths}} \times \underbrace{\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}}_{\text{term-to-concept similarity matrix}}$$

Example: the matrix contains 7 documents with the corresponding frequencies of each term.

In real IR applications, we take into considerations the normalized TF and IDF when calculating term weights.

The rank of this matrix r=2 because we have 2 types of documents (CS and Medical documents), i.e. 2 concepts.

SVD

$$
\begin{array}{c} \\ \text{CS} \\ \\ \\ \text{MD} \\ \\ \end{array}
\begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}
= \underbrace{\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}}_{\text{doc-to-concept similarity matrix}}
\text{x} \underbrace{\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}}_{\text{concepts strengths}}
\text{x} \underbrace{\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}}_{\text{term-to-concept similarity matrix}}
$$

U can be thought as the document-to-concept similarity matrix, while V is the term-to-concept similarity matrix.

For example, $U_{1,1}$ is the weight of CS concept in document d_1, λ_1 is the strength of the CS concept, $V_{1,1}$ is the weight of the first term 'data' in the CS concept, $V_{2,1}$=0 means 'data' has zero similarity with the 2nd concept (Medical).

What does $U_{4,1}$ means?

SVD

The N x N matrix $D = A \times A^T$ will intuitively give the document-to-document similarities.

$$\mathbf{D} = \mathbf{A} \times \mathbf{A}^t = \begin{bmatrix} 3 & 6 & 3 & 15 & 0 & 0 & 0 \\ 6 & 12 & 6 & 30 & 0 & 0 & 0 \\ 3 & 6 & 3 & 15 & 0 & 0 & 0 \\ 15 & 30 & 15 & 75 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 12 & 4 \\ 0 & 0 & 0 & 0 & 12 & 18 & 6 \\ 0 & 0 & 0 & 0 & 4 & 6 & 2 \end{bmatrix}$$

The eigenvectors of the D matrix will be the columns of the U matrix of the SVD of A. Proof?

$A \times A^T = U \times \Lambda \times V^T \times V \times \Lambda \times U^T$ (remember that $(A \times B)^T = B^T \times A^T$)

$= U \times \Lambda \times \Lambda \times U^T$: $V^T \times V = I$ because V is orthonormal

$= U \times \Lambda^2 \times U^T$: because Λ is a diagonal matrix.

Notice that D is symmetric because $(U \times \Lambda^2 \times U^T)^T = U \times \Lambda^2 \times U^T$

or simply because $(A \times A^T)^T = A^{TT} \times A^T = A \times A^T$

Because D is symmetric, it can be written as $D = S = U \times \Lambda` \times U^T$ where U are the D's eigenvectors and $\Lambda`$ are D's eigenvalues.

SVD

Symmetrically, the n x n matrix $T = A^T \times A$ will give the term-to-term similarities (the covariance matrix).

$$\mathbf{T} = \mathbf{A}^t \times \mathbf{A} \begin{bmatrix} 31 & 31 & 31 & 0 & 0 \\ 31 & 31 & 31 & 0 & 0 \\ 31 & 31 & 31 & 0 & 0 \\ 0 & 0 & 0 & 14 & 14 \\ 0 & 0 & 0 & 14 & 14 \end{bmatrix}$$

The eigenvectors of the T matrix are the columns of the V matrix of the SVD of A.

T is a symmetric matrix because $(A^T \times A)^T = A^T \times A^{TT} = A^T \times A$

$A^T \times A = V \times \Lambda \times U^T \times U \times \Lambda \times V^T = V \times \Lambda \times \Lambda \times V^T = V \times \Lambda^2 \times V^T$

Notice that both D and T have the same eigenvalues, which are the squares of the λ_i elements (The singular values of A).

These observations shows the close relation between SVD and PCA , which uses the eigenvectors of the covariance matrix.

SVD

Very important property: (we will see it again in Kleinberg algorithm)

$(A^T x\ A)^k\ x\ v` \approx$ (constant) $v_1$ where k>>1, v` is a random vector, v_1 is the first right singular vector of A, or equivalently is the eigenvector of A^T x A (as we already proved). *Proof*?

$(A^T\ x\ A)^k = (A^T\ x\ A)\ x\ (A^T\ x\ A)\ x \ldots = (V\ x\ \Lambda^2\ x\ V^T)\ x\ (V\ x\ \Lambda^2\ x\ V^t)\ x \ldots$

$= (V\ x\ \Lambda^2\ x\ V^T)\ x \ldots = (V\ x\ \Lambda^4\ x\ V^T)\ x \ldots. = (V\ x\ \Lambda^{2k}\ x\ V^T)$

Using spectral decomposition:

$(A^T x\ A)^k = (V\ x\ \Lambda^{2k}\ x\ V^T) = \lambda_1^{2k} v_1\ v_1^T + \lambda_2^{2k} v_2\ v_2^T + \ldots + \lambda_n^{2k} v_n\ v_n^T$

Because $\lambda_1 > \lambda_{i\neq1} \rightarrow \lambda_1^{2k} >> \lambda_{i\neq1}^{2k}$

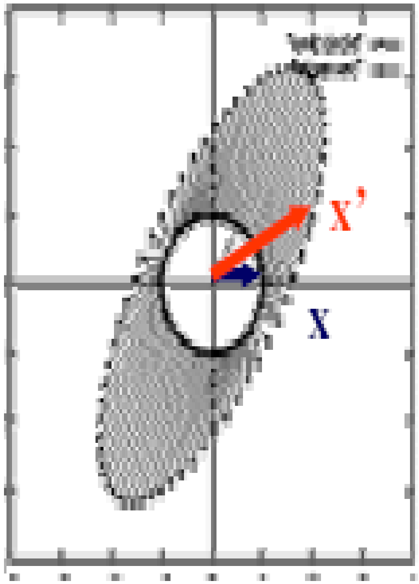
Thus $(A^T\ x\ A)^k \approx \lambda_1^{2k} v_1\ v_1^T$

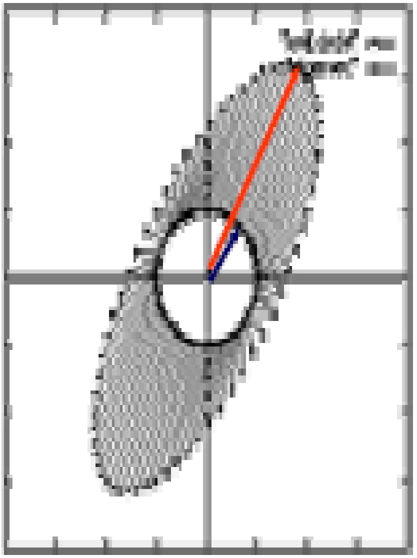
Now $(A^T x\ A)^k\ x\ v` = \lambda_1^{2k} v_1\ v_1^T\ x\ v` =$ (const) v_1

because $v_1^T x\ v`$ is a scalar.

SVD

Geometrically, this means that if we multiple any vector with matrix $(A^T \times A)^{k}$, then result is a vector that is parallel to the first eigenvector.





PCA and SVD

Summary for PCA and SVD

Objective: find the principal components P of a data matrix A(n,m).

1. First zero mean the columns of A (translate the origin to the center of gravity).
2. Apply PCA or SVD to find the principle components (P) of A.

PCA:

I. Calculate the covariance matrix C= $\frac{A\,A^T}{n}$

II. p = the eigenvectors of C.

III. The variances in each new dimension is given by the eigenvalues.

SVD:

I. Calculate the SVD of A.

II. P = V: the right singular vectors.

III. The variances are given by the squaring the singular values.

3. Project the data onto the feature space. F = P x A
4. Optional: Reconstruct A' from Y where A' is the compressed version of A.

Outline

- Mathematical background
- PCA
- SVD
- Some PCA and SVD applications
- Case study: LSI

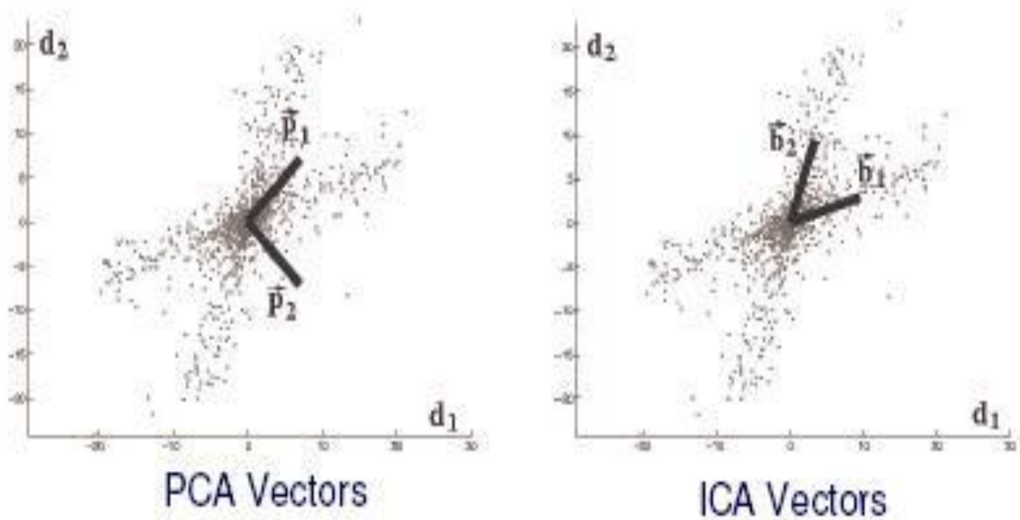
SVD and PCA applications

- LSI: Latent Semantic Indexing.
- Solve over-specified (no solution: least squares error solution) and under-specified (infinite number of solutions: shortest length solution) linear equations.
- Ratio rules (computer quantifiable association rules like *bread:milk:buffer=2:4:3*).
- Google/PageRank algorithm (random walk with restart).
- Kleinberg/Hits algorithm (compute hubs and authority scores for nodes).
- Query feedbacks (learn to estimate the selectivity of the queries: a regression problem).
- Image compression (*other methods: DCT used in JPEG, and wavelet compression*)
- Data visualization (by projecting the data on 2D).

Variations: ICA

ICA (Independent Components Analysis)

Relaxes the constraint of orthogonality but keeps the linearity. Thus, could be more flexible than PCA in finding patterns.



X (N,n)=H (N,n) x B (n,n)

where X is the data set, H are hidden variables, and B are basis vectors. $h_{i,j}$ can be understood as the weight of b_j in the instance X_i

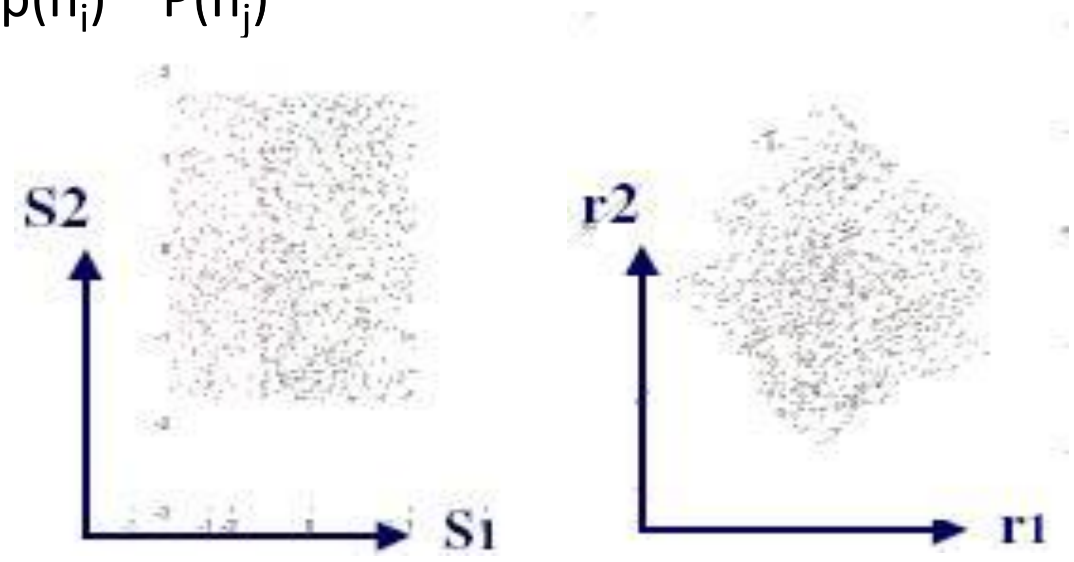
Variations: ICA

Linearity: $X_i = h_{i,1}\, b_1 + h_{i,2}\, b_2$

Problem definition: Knowing X, find H and B.

Make hidden variables h_i mutually independent:

$p(h_i, h_j) = p(h_i) * P(h_j)$



Which figure satisfies data independency?

Outline

- Mathematical background
- PCA
- SVD
- Some PCA and SVD applications
- Case study: LSI

SVD and LSI

LSI: Latent Semantic Indexing.

Idea: try to group similar terms together, to form a few concepts, then map the documents into vectors in the concept-space, as opposed to vectors in the n-dimensional space, where n is the vocabulary size of the document collection.

This approach automatically groups terms that occur together into concepts. Then every time the user asks for a term, the system determines the relevant concepts and search for them.

In order to map documents or queries into the concept space, we need the term-to-concept similarity matrix V.

SVD and LSI

Example: find the documents containing the term 'data'.

$$\vec{q} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

To translate q to a vector q_c in the concept space:

$q_c = V^T \times q$

$$= \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.58 \\ 0 \end{bmatrix}$$

It means that q is related to the CS group of terms (with strength=0.58), and unrelated to the medical group of terms.

SVD and LSI

More importantly, q_c now involves the terms ‘information’ and ‘retrieval’,

$$\vec{q_c} = \begin{bmatrix} 0.58 \\ 0 \end{bmatrix}$$

thus LSI system may return documents that do not necessarily contain the term ‘data’

For example, a document d with a single word ‘retrieval’ $\vec{d} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$

d will be mapped into the concept space $\vec{d_c} = V^t \times \vec{d_c} = \begin{bmatrix} 0.58 \\ 0 \end{bmatrix} = \vec{q_c}$

And will be a perfect match for the query.

Cosine similarity is one way to measure the similarity between the query and the documents.

Experiments showed that LSI outperforms standard vector methods with improvement of as much as 30% in terms of precision and recall.

Thank you for listening

Questions or Thoughts??

Dan Jurafsky and James Martin

Speech and Language Processing

Chapter 6:
Vector Semantics

What do words mean?

First thought: look in a dictionary

http://www.oed.com/

Words, Lemmas, Senses, Definitions

lemma

sense

definition

pepper, *n.*

Pronunciation: Brit. /ˈpɛpə/ , U.S. /ˈpɛpər/

Forms: OE **peopor** (*rare*), OE **pipcer** (transmission error), OE **pipor**, OE **pipur** (*rare* .

Frequency (in current use):

Etymology: A borrowing from Latin. **Etymon:** Latin *piper*.

< classical Latin *piper*, a loanword < Indo-Aryan (as is ancient Greek *πέπερι*); compare Sar

I. The spice or the plant.

1.

a. A hot pungent spice derived from the prepared fruits (peppercorns) of the pepper plant, *Piper nigrum* (see sense 2a), used from early times to season food, either whole or ground to powder (often in association with salt). Also (locally, chiefly with distinguishing word): a similar spice derived from the fruits of certain other species of the genus *Piper*; the fruits themselves.

The ground spice from *Piper nigrum* comes in two forms, the more pungent *black pepper*, produced from black peppercorns, and the milder *white pepper*, produced from white peppercorns: see BLACK *adj.* and *n.* Special uses 5a, PEPPERCORN *n.* 1a, and WHITE *adj.* and *n.*[1] Special uses 7b(a).

2.

a. The plant *Piper nigrum* (family Piperaceae), a climbing shrub indigenous to South Asia and also cultivated elsewhere in the tropics, which has alternate stalked entire leaves, with pendulous spikes of small green flowers opposite the leaves, succeeded by small berries turning red when ripe. Also more widely: any plant of the genus *Piper* or the family Piperaceae.

b. Usu. with distinguishing word: any of numerous plants of other families having hot pungent fruits or leaves which resemble pepper (1a) in taste and in some cases are used as a substitute for it.

c. *U.S.* The California pepper tree, *Schinus molle*. Cf. PEPPER TREE *n.*

3. Any of various forms of capsicum, esp. *Capsicum annuum* var. *annuum*. Originally (chiefly with distinguishing word): any variety of the *C. annuum* Longum group, with elongated fruits having a hot, pungent taste, the source of cayenne, chilli powder, paprika, etc., or of the perennial *C. frutescens*, the source of Tabasco sauce. Now frequently (more fully ***sweet pepper***): any variety of the *C. annuum* Grossum group, with large, bell-shaped or apple-shaped, mild-flavoured fruits, usually ripening to red, orange, or yellow and eaten raw in salads or cooked as a vegetable. Also: the fruit of any of these capsicums.

Sweet peppers are often used in their green immature state (more fully ***green pepper***), but some new varieties remain green when ripe.

Lemma pepper

Sense 1: spice from pepper plant

Sense 2: the pepper plant itself

Sense 3: another similar plant (Jamaican pepper)

Sense 4: another plant with peppercorns (California pepper)

Sense 5: *capsicum* (i.e. chili, paprika, bell pepper, etc)

A sense or “concept” is the meaning component of a word

There are relations between senses

Relation: Synonymity

Synonyms have the same meaning in some or all contexts.

- filbert / hazelnut
- couch / sofa
- big / large
- automobile / car
- vomit / throw up
- Water / H_20

Relation: Synonymity

Note that there are probably no examples of perfect synonymy.

- Even if many aspects of meaning are identical
- Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

The Linguistic Principle of Contrast:

- Difference in form -> difference in meaning

Relation: Synonymity?

Water/H_20

Big/large

Brave/courageous

Relation: Antonymy

Senses that are opposites with respect to one feature of meaning

Otherwise, they are very similar!

`dark/light` `short/long` `fast/slow` `rise/fal`

`hot/cold` `up/down` `in/out`

More formally: antonyms can

- define a binary opposition
  or be at opposite ends of a scale
  - `long/short, fast/slow`
- Be *reversives*:
  - `rise/fall, up/down`

Relation: Similarity

Words with similar meanings. Not synonyms, but sharing some element of meaning

```
car, bicycle
cow, horse
```

Ask humans how similar 2 words are

word1	word2	similarity
vanish	disappear	9.8
behave	obey	7.3
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

SimLex-999 dataset (Hill et al., 2015)

Relation: Word relatedness

Also called "word association"

Words be related in any way, perhaps via a semantic frame or field

- `car, bicycle`: **similar**
- `car, gasoline`: **related**, not similar

Semantic field

Words that

- cover a particular semantic domain
- bear structured relations with each other.

hospitals

surgeon, *scalpel*, *nurse*, *anaesthetic*, *hospital*

restaurants

waiter, menu, plate, food, menu, chef),

houses

door, roof, kitchen, family, bed

Relation: Superordinate/ subordinate

One sense is a **subordinate** of another if the first sense is more specific, denoting a subclass of the other

- *car* is a subordinate of *vehicle*
- *mango* is a subordinate of *fruit*

Conversely **superordinate**

- *vehicle* is a superordinate of *car*
- *fruit* is a subodinate of *mango*

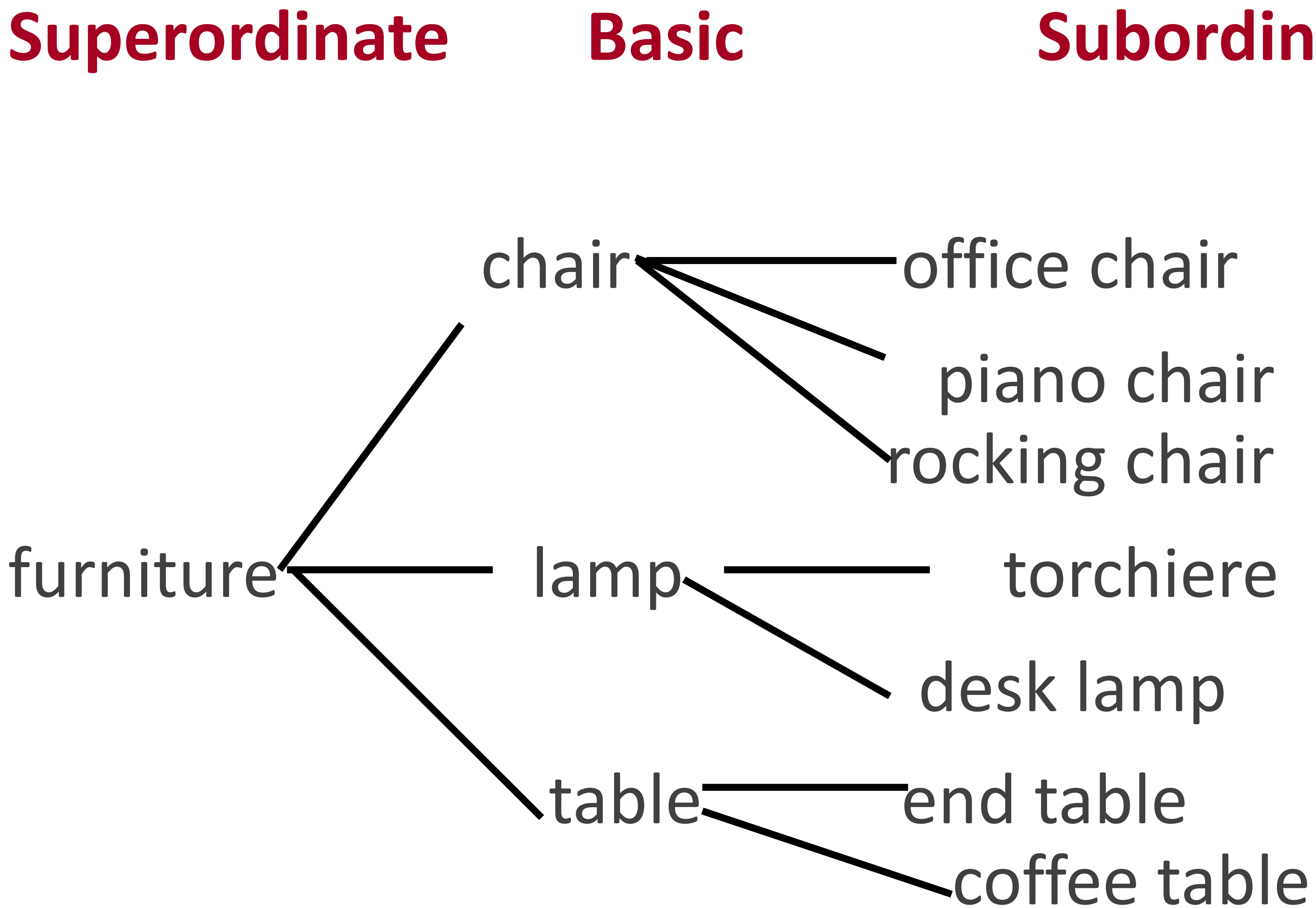
Superordinate	vehicle	fruit	furniture
Subordinate	car	mango	chair

These levels are not symmetric

One level of category is distinguished from the others

The "basic level"

Name these items


Superordinate
Basic
Subordinate
furniture
chair
office chair
piano chair
rocking chair
lamp
torchiere
desk lamp
table
end table
coffee table

Cluster of Interactional Properties

Basic level things are “human-sized”

Consider chairs

- We know how to interact with a chair (sitting)
- Not so clear for superordinate categories like furniture
  - “Imagine a furniture without thinking of a bed/table/chair/specific basic-level category”

The basic level

Is the level of distinctive actions

Is the level which is learned earliest and at which things are first named

It is the level at which names are shortest and used most frequently

Connotation

Words have **affective** meanings

positive connotations (*happy*)

negative connotations (*sad*)

positive evaluation (*great*, *love*)

negative evaluation (*terrible*, *hate*).

So far

Concepts or word senses

- Have a complex many-to-many association with **words** (homonymy, multiple senses)

Have relations with each other

- Synonymy
- Antonymy
- Similarity
- Relatedness
- Superordinate/subordinate
- Connotation

But how to define a concept?

Classical (“Aristotelian”) Theory of Concepts

The meaning of a word:

a concept defined by **necessary** and **sufficient** conditions

A **necessary** condition for being an X is a condition C that X must satisfy in order for it to be an X.

- If not C, then not X
- ”Having four sides” is necessary to be a square.

A **sufficient** condition for being an X is condition such that if something satisfies condition C, then it must be an X.

- If and only if C, then X
- The following necessary conditions, jointly, are sufficient to be a square
  - x has (exactly) four sides
  - each of x's sides is straight
  - x is a closed figure
  - x lies in a plane
  - each of x's sides is equal in length to each of the others
  - each of x's interior angles is equal to the others (right angles)
  - the sides of x are joined at their ends

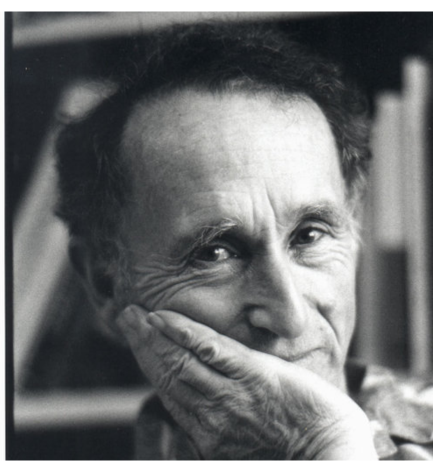
Example from Norman Swartz, SFU

Problem 1: The features are complex and may be context-dependent

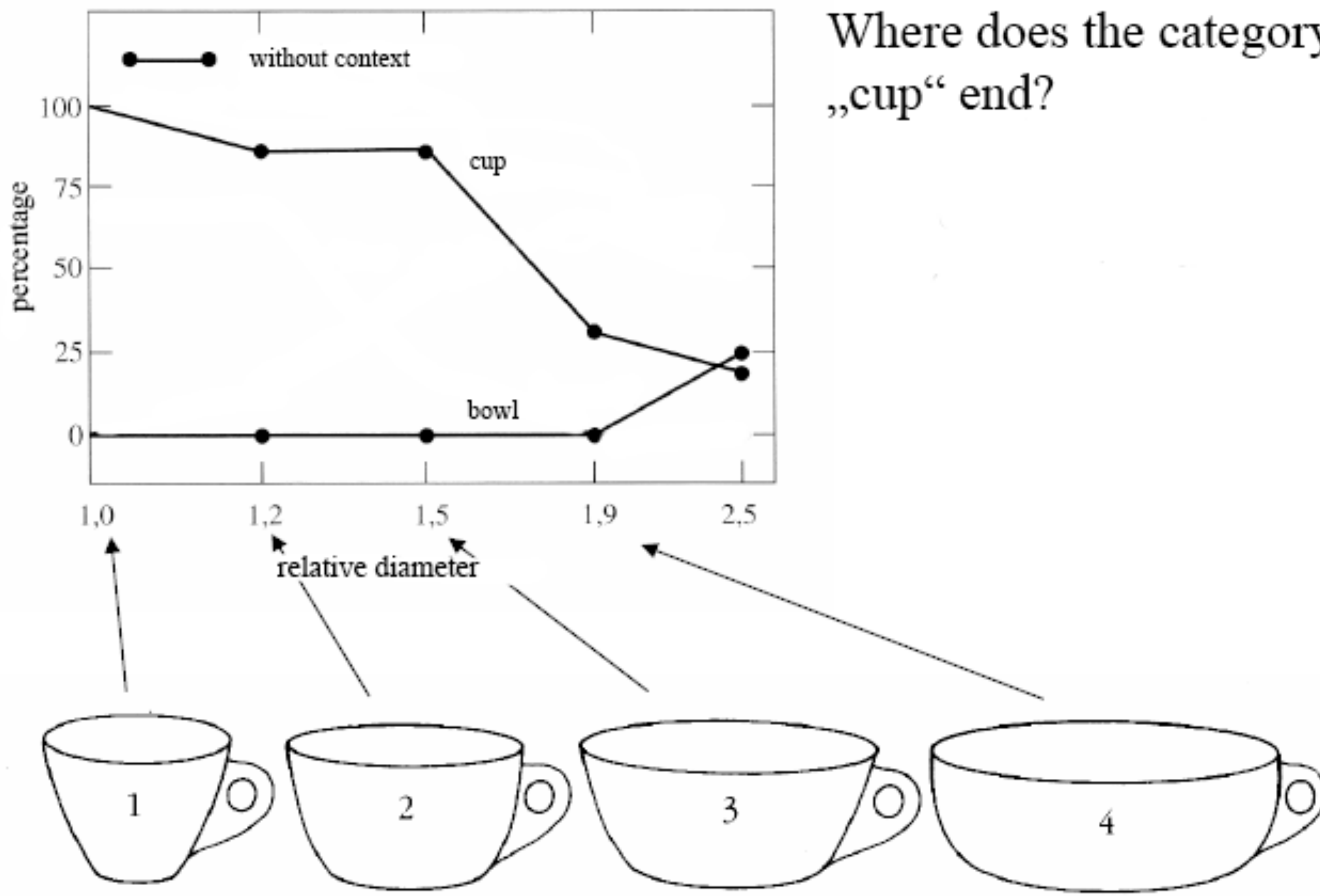
William Labov. 1975

What are these?

Cup or bowl?

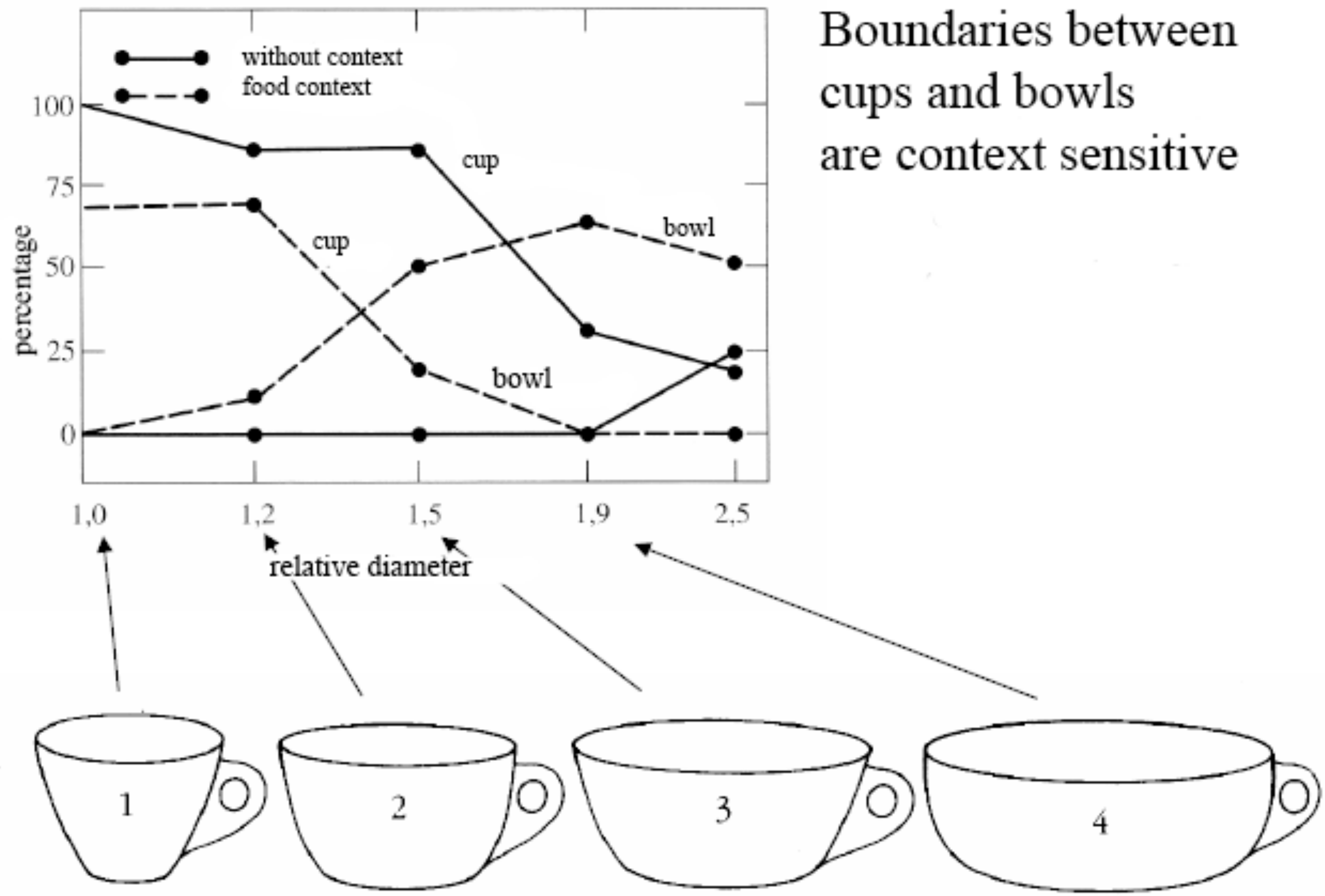


The category depends on complex features of the object (diameter, etc)



Where does the category „cup“ end?

The category depends on the context! (If there is food in it, it's a bowl)



Boundaries between cups and bowls are context sensitive

Labov's definition of cup

The term *cup* is used to denote round containers with a ratio of depth to width of 1±r where $r \leq r_b$, and $r_b = \alpha_1 + \alpha_2 + \ldots \alpha_\upsilon$ and α_1 is a positive quality when the feature i is present and 0 otherwise.

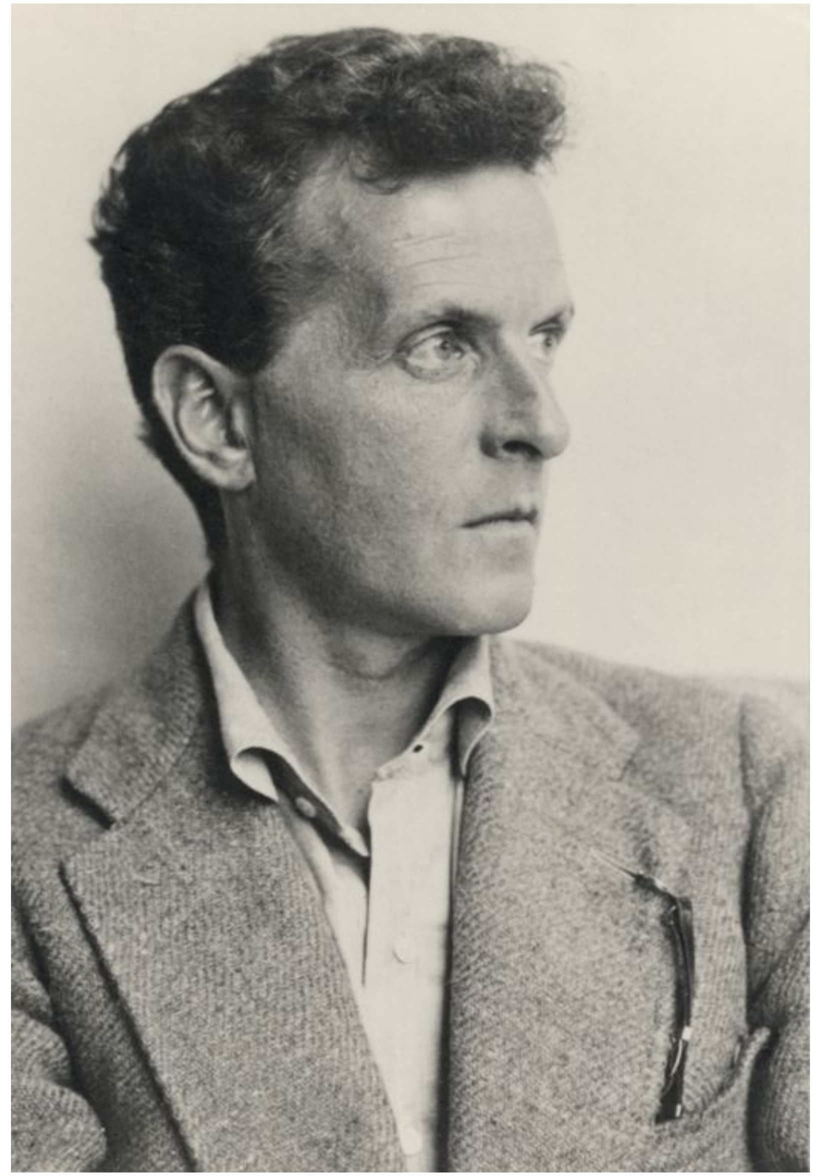
feature 1 = with one handle
2 = made of opaque vitreous material
3 = used for consumption of food
4 = used for the consumption of liquid food
5 = used for consumption of hot liquid food
6 = with a saucer
7 = tapering
8 = circular in cross-section

Cup is used variably to denote such containers with ratios width to depth 1±r where $r_b \leq r \leq r_1$ with a probability of r_1 - r/r_t – r_b. The quantity $1 \pm r_b$ expresses the distance from the modal value of width to height.

Ludwig Wittgenstein (1889-1951)

Philosopher of language

In his late years, a proponent of studying "ordinary language"



Wittgenstein (1945) *Philosophical Investigations.* Paragraphs 66,67

66. Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There *must* be something common, or they would not be called 'games' "—but *look and see* whether there is anything common to all.—For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!—Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.—Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.

And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.

67. I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way.—And I shall say: 'games' form a family.

And for instance the kinds of number form a family in the same way. Why do we call something a "number"? Well, perhaps because it has a—direct—relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.

But if someone wished to say: "There is something common to all these constructions—namely the disjunction of all their common properties"—I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread—namely the continuous overlapping of those fibres".

What is a game?

Wittgenstein’s thought experiment on "What is a game”:

PI #66:

”Don’t say “there must be something common, or they would not be called `games’”—but *look and see* whether there is anything common to all”

Is it amusing?

Is there competition?

Is there long-term strategy?

Is skill required?

Must luck play a role?

Are there cards?

Is there a ball?

Family Resemblance

Game 1	Game 2	Game 3	Game 4
ABC	BCD	ACD	ABD

“each item has at least one, and probably several, elements in common with one or more items, but no, or few, elements are common to all items”   Rosch and Mervis

How about a radically different approach?

Ludwig Wittgenstein

PI #43:

"The meaning of a word is its use in the language"

Let's define words by their usages

In particular, words are defined by their environments (the words around them)

Zellig Harris (1954): **If A and B have almost identical environments we say that they are synonyms**.

What does ongchoi mean?

Suppose you see these sentences:

- Ong choi is delicious **sautéed with garlic**.
- Ong choi is superb **over rice**
- Ong choi **leaves** with salty sauces

And you've also seen these:

- …spinach **sautéed with garlic over rice**
- Chard stems and **leaves** are **delicious**
- Collard greens and other **salty** leafy greens

Conclusion:

- Ongchoi is a leafy green like spinach, chard, or collard greens

Ong choi: *Ipomoea aquatica* *"Water Spinach"*

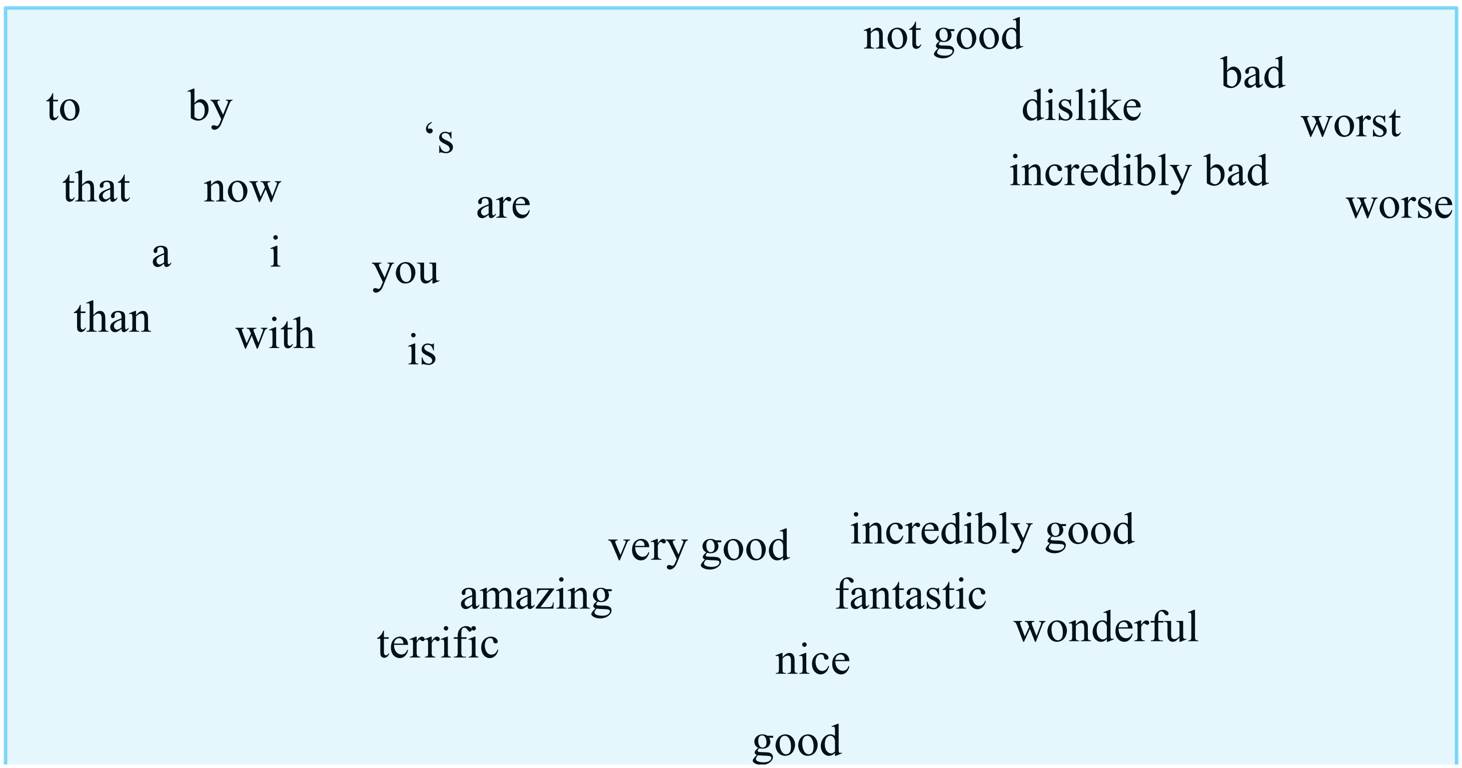
Yamaguchi, Wikimedia Commons, public domain

We'll build a new model of meaning focusing on similarity

Each word = a vector

- Not just "word" or word45.

Similar words are "nearby in space"



We define a word as a vector

Called an "embedding" because it's embedded into a space

The standard way to represent meaning in NLP

Fine-grained model of meaning for similarity

- NLP tasks like sentiment analysis
  - With words, requires **same** word to be in training and test
  - With embeddings: ok if **similar** words occurred!!!
- Question answering, conversational agents, etc

We'll introduce 2 kinds of embeddings

Tf-idf

- A common baseline model
- Sparse vectors
- Words are represented by a simple function of the counts of nearby words

Word2vec

- Dense vectors
- Representation is created by training a classifier to distinguish nearby and far-away words

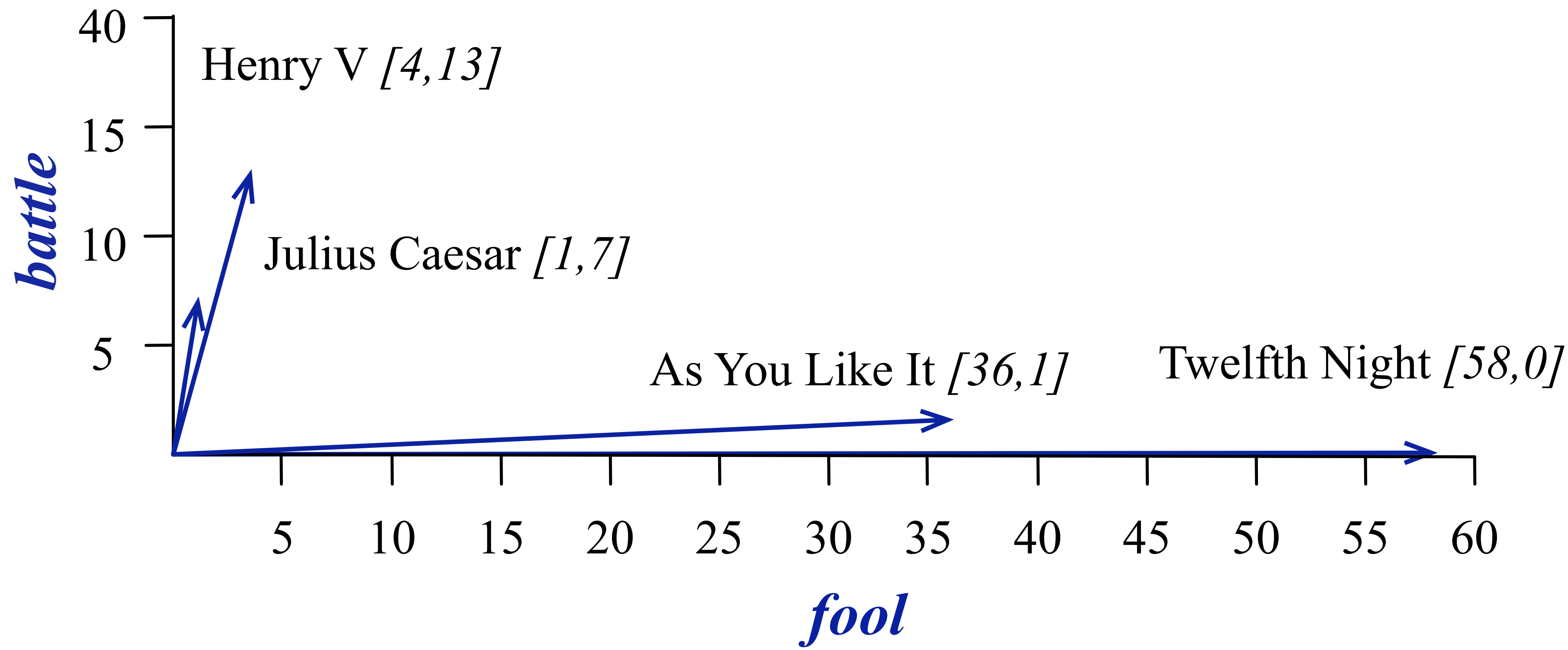
Review: words, vectors, and co-occurrence matrices

Term-document matrix

Each document is represented by a vector of words

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Visualizing document vectors



Vectors are the basis of information retrieval

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Vectors are similar for the two comedies
Different than the history

Comedies have more fools and wit and fewer battles.

Words can be vectors too

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

battle is "the kind of word that occurs in Julius Caesar and Henry V"

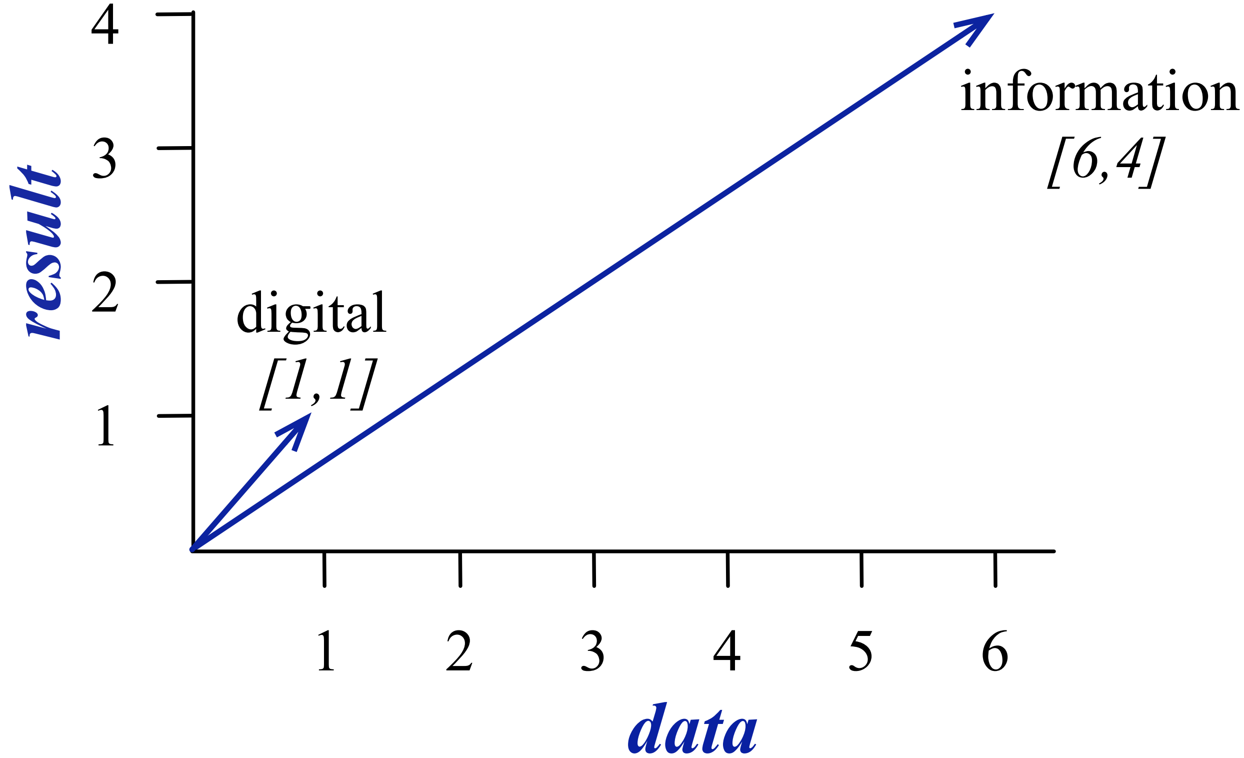
fool is "the kind of word that occurs in comedies, especially Twelfth Night"

More common: word-word matrix (or "term-context matrix")

Two **words** are similar in meaning if their context vectors are similar

sugar, a sliced lemon, a tablespoonful of	**apricot**	jam, a pinch each of,
their enjoyment. Cautiously she sampled her first	**pineapple**	and another fruit whose taste she likened
well suited to programming on the digital	**computer**.	In finding the optimal R-stage policy from
for the purpose of gathering data and	**information**	necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	…
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	


4
3
2
1
result
digital
[1,1]
information
[6,4]
1
2
3
4
5
6
data

Reminders from linear algebra

$$\text{dot-product}(\vec{v},\vec{w}) = \vec{v}\cdot\vec{w} = \sum_{i=1}^{N} v_i w_i = v_1 w_1 + v_2 w_2 + \ldots + v_N w_N$$

vector length $|\vec{v}| = \sqrt{\sum_{i=1}^{N} v_i^2}$

Cosine for computing similarity

Sec. 6.3

$$\text{cosine}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

v_i is the count for word v in context i

w_i is the count for word w in context i.

$$\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta$$

$$\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} = \cos\theta$$

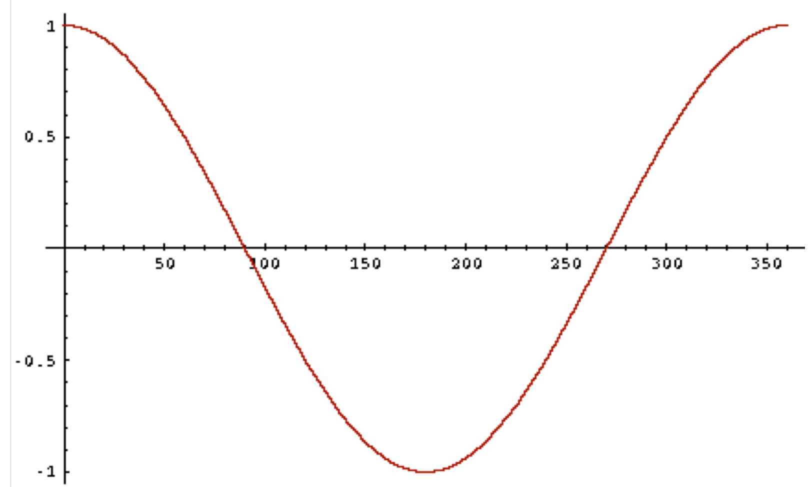
Cos($\vec{v}$,$\vec{w}$) is the cosine similarity of $\vec{v}$ and $\vec{w}$

Cosine as a similarity metric

-1: vectors point in opposite directions

+1: vectors point in same directions

0: vectors are orthogonal



Frequency is non-negative, so cosine range 0-1

$$\cos(\vec{v},\vec{w}) = \frac{\vec{v}\bullet\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\vec{v}}{|\vec{v}|}\bullet\frac{\vec{w}}{|\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

	large	data	computer
apricot	1	0	0
digital	0	1	2
information	1	6	1

Which pair of words is more similar?

cosine(apricot,information) =

$$\frac{1+0+0}{\sqrt{1+0+0}\ \sqrt{1+36+1}} = \frac{1}{\sqrt{38}} = .16$$

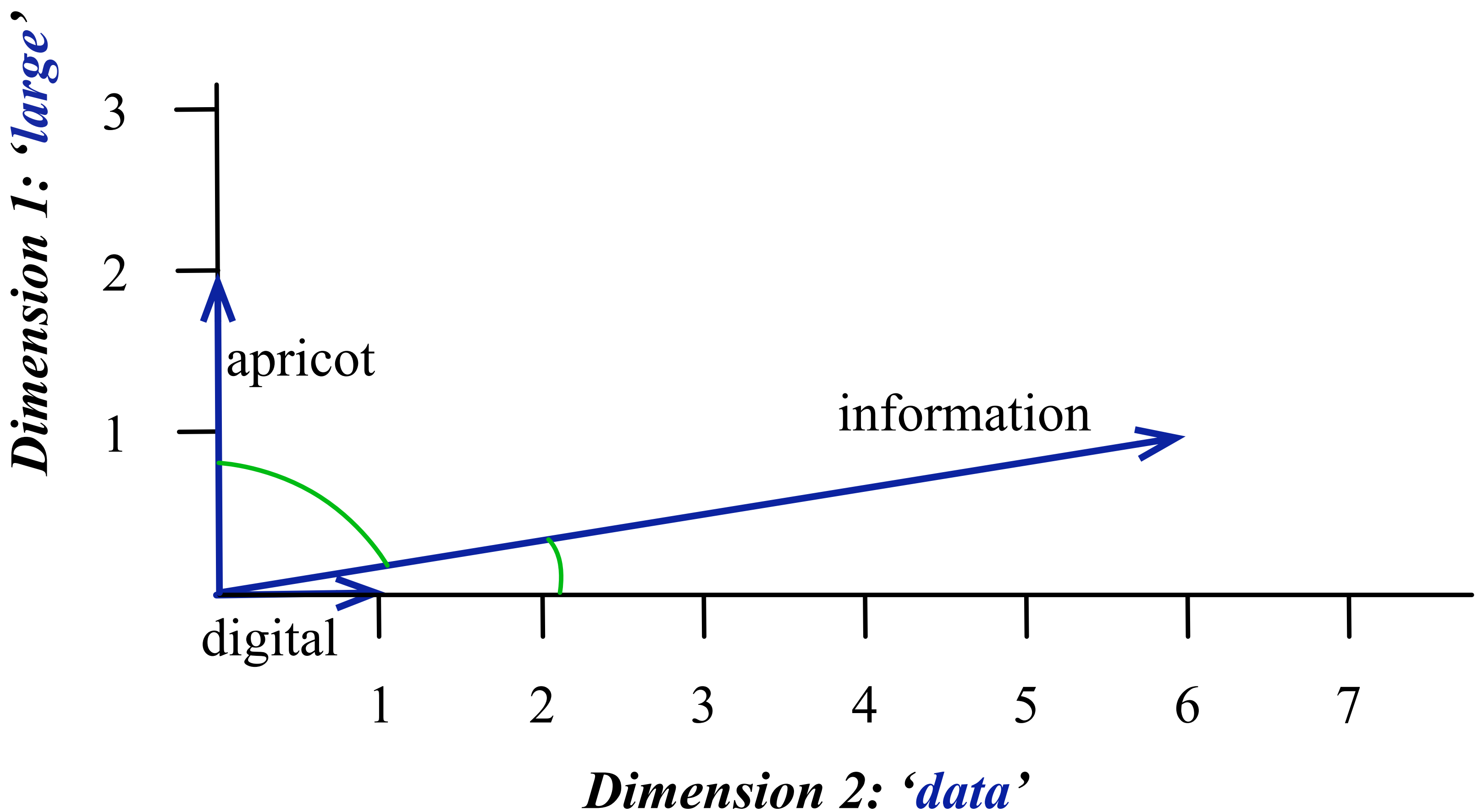
cosine(digital,information) =

$$\frac{0+6+2}{\sqrt{0+1+4}\ \sqrt{1+36+1}} = \frac{8}{\sqrt{38}\sqrt{5}} = .58$$

cosine(apricot,digital) =

$$\frac{0+0+0}{\sqrt{1+0+0}\ \sqrt{0+1+4}} = 0$$

Visualizing cosines (well, angles)



But raw frequency is a bad representation

Frequency is clearly useful; if *sugar* appears a lot near *apricot*, that's useful information.

But overly frequent words like *the*, *it,* or *they* are not very informative about the context

Need a function that resolves this frequency paradox!

tf-idf: combine two factors

tf: term frequency. frequency count (usually log-transformed):

$$\mathrm{tf}_{t,d} = \begin{cases} 1 + \log_{10} \mathrm{count}(t,d) & \text{if } \mathrm{count}(t,d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Idf: inverse document frequency: tf-

Total # of docs in collection

$$\mathrm{idf}_i = \log\left(\frac{N}{\mathrm{df}_i}\right)$$

Words like "the" or "good" have very low idf

# of docs that have word i

tf-idf value for word t in document d:

$$w_{t,d} = \mathrm{tf}_{t,d} \times \mathrm{idf}_t$$

Summary: tf-idf

Compare two words using tf-idf cosine to see if they are similar

Compare two documents

- Take the centroid of vectors of all the words in the document
- Centroid document vector is:

$$d = \frac{w_1 + w_2 + ... + w_k}{k}$$

An alternative to tf-idf

Ask whether a context word is **particularly informative** about the target word.

- Positive Pointwise Mutual Information (PPMI)

Pointwise Mutual Information

Pointwise mutual information:

Do events x and y co-occur more than if they were independent?

$$\mathrm{PMI}(X,Y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

PMI between two words: (Church & Hanks 1989)

Do words x and y co-occur more than if they were independent?

$$\mathrm{PMI}(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

Positive Pointwise Mutual Information

- PMI ranges from $-\infty$ to $+\infty$
- But the negative values are problematic
  - Things are co-occurring **less than** we expect by chance
  - Unreliable without enormous corpora
    - Imagine w1 and w2 whose probability is each 10^{-6}
    - Hard to be sure p(w1,w2) is significantly different than 10^{-12}
  - Plus it's not clear people are good at "unrelatedness"
- So we just replace negative PMI values by 0
- Positive PMI (PPMI) between word1 and word2:

$$\mathrm{PPMI}(word_1, word_2) = \max\left(\log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}, 0\right)$$

Computing PPMI on a term-context matrix

Matrix F with W rows (words) and C columns (contexts)

f_{ij} is # of times w_i occurs in context c_j

	aardvark	computer	data	pinch	result	sugar
apricot	0	0	0	1	0	1
pineapple	0	0	0	1	0	1
digital	0	2	1	0	1	0
information	0	1	6	0	4	0

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}} \qquad p_{i*} = \frac{\sum_{j=1}^{C} f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}} \qquad p_{*j} = \frac{\sum_{i=1}^{W} f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}}$$

$$pmi_{ij} = \log_2 \frac{p_{ij}}{p_{i*}p_{*j}} \qquad ppmi_{ij} = \begin{cases} pmi_{ij} & \text{if } pmi_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}}$$

Count(w,context)

	computer	data	pinch	result	sugar
apricot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	1	0
information	1	6	0	4	0

p(w=information,c=data) = 6/19 = .32

p(w=information) = 11/19 = .58

p(c=data) = 7/19 = .37

$$p(w_i) = \frac{\sum_{j=1}^{C} f_{ij}}{N} \qquad p(c_j) = \frac{\sum_{i=1}^{W} f_{ij}}{N}$$

p(w,context) **p(w)**

	computer	data	pinch	result	sugar	p(w)
apricot	0.00	0.00	0.05	0.00	0.05	0.11
pineapple	0.00	0.00	0.05	0.00	0.05	0.11
digital	0.11	0.05	0.00	0.05	0.00	0.21
information	0.05	0.32	0.00	0.21	0.00	0.58
p(context)	0.16	0.37	0.11	0.26	0.11	

$$pmi_{ij} = \log_2 \frac{p_{ij}}{p_{i*}p_{*j}}$$

	p(w,context)					p(w)
	computer	data	pinch	result	sugar	
apricot	0.00	0.00	0.05	0.00	0.05	0.11
pineapple	0.00	0.00	0.05	0.00	0.05	0.11
digital	0.11	0.05	0.00	0.05	0.00	0.21
information	0.05	0.32	0.00	0.21	0.00	0.58
p(context)	0.16	0.37	0.11	0.26	0.11	

pmi(information,data) = $\log_2$ (.32 / (.37*.58)) = .58

(.57 using full precision)

PPMI(w,context)

	computer	data	pinch	result	sugar
apricot	-	-	2.25	-	2.25
pineapple	-	-	2.25	-	2.25
digital	1.66	0.00	-	0.00	-
information	0.00	0.57	-	0.47	-

Weighting PMI

PMI is biased toward infrequent events

- Very rare words have very high PMI values

Two solutions:

- Give rare words slightly higher probabilities
- Use add-one smoothing (which has a similar effect)

Weighting PMI: Giving rare context words slightly higher probability

Raise the context probabilities to $\alpha = 0.75$:

$$\text{PPMI}_\alpha(w,c) = \max\left(\log_2 \frac{P(w,c)}{P(w)P_\alpha(c)}, 0\right)$$

$$P_\alpha(c) = \frac{count(c)^\alpha}{\sum_c count(c)^\alpha}$$

This helps because $P_\alpha(c) > P(c)$ for rare c

Consider two events, P(a) = .99 and P(b)=.01

$$P_\alpha(a) = \frac{.99^{.75}}{.99^{.75}+.01^{.75}} = .97 \quad P_\alpha(b) = \frac{.01^{.75}}{.01^{.75}+.01^{.75}} = .03$$

Use Laplace (add-1) smoothing

Add-2 Smoothed Count(w,context)

	computer	data	pinch	result	sugar
apricot	2	2	3	2	3
pineapple	2	2	3	2	3
digital	4	3	2	3	2
information	3	8	2	6	2

p(w,context) [add-2] **p(w)**

	computer	data	pinch	result	sugar	p(w)
apricot	0.03	0.03	0.05	0.03	0.05	0.20
pineapple	0.03	0.03	0.05	0.03	0.05	0.20
digital	0.07	0.05	0.03	0.05	0.03	0.24
information	0.05	0.14	0.03	0.10	0.03	0.36
p(context)	0.19	0.25	0.17	0.22	0.17	

PPMI versus add-2 smoothed PPMI

PPMI(w,context)

	computer	data	pinch	result	sugar
apricot	-	-	2.25	-	2.25
pineapple	-	-	2.25	-	2.25
digital	1.66	0.00	-	0.00	-
information	0.00	0.57	-	0.47	-

PPMI(w,context) [add-2]

	computer	data	pinch	result	sugar
apricot	0.00	0.00	0.56	0.00	0.56
pineapple	0.00	0.00	0.56	0.00	0.56
digital	0.62	0.00	0.00	0.00	0.00
information	0.00	0.58	0.00	0.37	0.00

Summary for Part I

- Survey of Lexical Semantics
- Idea of Embeddings: Represent a word as a function of its distribution with other words
- Tf-idf
- Cosines
- PPMI

- Next lecture: sparse embeddings, word2vec