CHAPTER 2

Determinants and Eigenvalues

1. Introduction

Gauss-Jordan reduction is an extremely effective method for solving systems of linear equations, but there are some important cases in which it doesn't work very well. This is particularly true if some of the matrix entries involve symbolic parameters rather than specific numbers.

EXAMPLE 1.1. Solve the general 2×2 system

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

We can't easily use Gaussian reduction since it would proceed differently if a were zero than it would if a were not zero. In this particular case, however, the usual high school method works quite well. Multiply the first equation by d and the second by b to obtain

$$\begin{aligned} adx + bdy &= ed \\ bcx + bdy &= bf \end{aligned}$$

and subtract to obtain

$$\begin{aligned} adx - bcx &= ed - bf \\ \text{or} \qquad (ad - bc)x &= ed - bf \\ \text{or} \qquad x &= \frac{ed - bf}{ad - bc}. \end{aligned}$$

Similarly, multiplying the first equation by c, the second by a and subtracting yields

$$y = \frac{af - ce}{ad - bc}.$$

For this to work, we must assume only that $ad - bc \neq 0$.

(See the exercises for a slightly different but equivalent approach which uses A^{-1}.)

You should recognize the quantity in the denominator as the 2×2 *determinant*

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

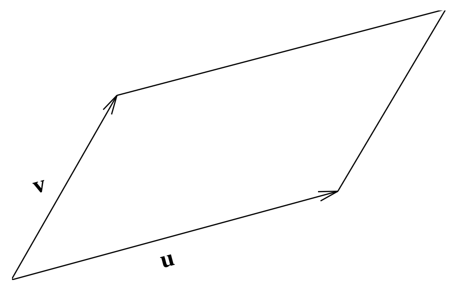
2×2 determinants arise in a variety of important situations. For example, if

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

are two vectors in the plane, then

$$\det \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} = u_1 v_2 - v_1 u_2$$

is the *signed area* of the parallelogram spanned by the vectors.

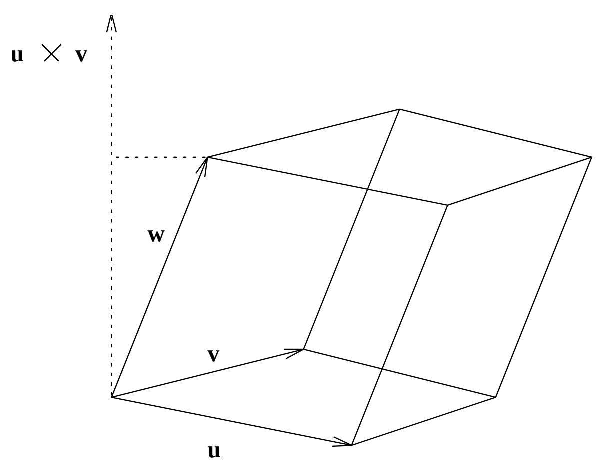


The sign is determined by the orientation of the two vectors. It is positive if the smaller of the two angles from $\mathbf{u}$ to $\mathbf{v}$ is in the counterclockwise direction and negative if it is in the clockwise direction.

You are probably also familiar with 3×3 determinants. For example, if $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in $\mathbf{R}^3$, the triple product

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$$

gives the *signed volume* of the parallelepiped spanned by the three vectors.



The sign is positive if the vectors form a 'right handed' triple and it is negative if they form a 'left handed triple'. If you express this triple product in terms of components, you obtain the expression

$$u_1 v_2 w_3 + v_1 w_2 u_3 + w_1 u_2 v_3 - u_1 w_2 v_3 - v_1 w_2 u_3 - w_1 v_2 u_3$$

and this quantity is called the determinant of the 3×3 matrix

$$\begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}.$$

As you might expect, if you try to solve the general 3×3 system

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

without having specific numerical values for the entries of the coefficient matrix, then you end up with a bunch of formulas in which the 3×3 determinant of the coefficient matrix A plays a significant role.

Our program in this chapter will be to generalize these concepts to arbitrary $n \times n$ matrices. This is necessarily a bit complicated because of the complexity of the formulas. For $n > 3$, it is not easy to write out explicit formulas for the quantities which arise, so we use another approach.

Exercises for Section 1.

1.1. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$. Verify that $\pm|\mathbf{u} \times \mathbf{v}| = \det \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$. This in effect shows that, except for sign, a 2×2 determinant is the area of the parallelogram spanned by its columns in $\mathbf{R}^2$.

1.2. Let

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(a) Calculate the determinants of the following 2×2 matrices

$$(i)\ \begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}, \qquad (ii)\ \begin{bmatrix} \mathbf{v} & \mathbf{u} \end{bmatrix}, \qquad (iii)\ \begin{bmatrix} \mathbf{u} - 2\mathbf{v} & \mathbf{v} \end{bmatrix}.$$

(b) Draw plane diagrams for each of the parallelograms spanned by the columns of these matrices. Explain geometrically what happened to the area.

1.3. Let $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ be the columns of the matrix

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

(a) Find $\det A$ by computing $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ and check by computing $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

(b) Without doing the computation, find $\det \begin{bmatrix} \mathbf{v} & \mathbf{u} & \mathbf{w} \end{bmatrix}$, $\det \begin{bmatrix} \mathbf{u} & \mathbf{w} & \mathbf{v} \end{bmatrix}$, and $\det \begin{bmatrix} \mathbf{w} & \mathbf{v} & \mathbf{u} \end{bmatrix}$.

(c) Explain why the determinant of the above matrix does not change if you replace the first column by the the sum of the first two columns.

(d) What happens if you multiply one of the columns by -3?

1.4. Solve the system

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}$$

by multiplying the right hand side by the inverse of the coefficient matrix. Compare what you get with the solution obtained in the section.

2. Definition of the Determinant

Let A be an $n \times n$ matrix. By definition

$$\text{for } n = 1 \qquad \det \begin{bmatrix} a \end{bmatrix} = a \tag{9}$$

$$\text{for } n = 2 \qquad \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{10}$$

As mentioned in the previous section, we can give an explicit formula to define $\det A$ for $n = 3$, but an explicit formula for larger n is very difficult to describe. Here is a provisional definition. Form a sum of many terms as follows. Choose any entry from the first row of A; there are n possible ways to do that. Next, choose any entry from the second row which is not in the same column as the first entry chosen; there are $n - 1$ possible ways to do that. Continue in this way until you have chosen one entry from each row in such a way that no column is repeated; there are $n!$ ways to do that. Now multiply all these entries together to form a typical term. If that were all, it would be complicated enough, but there is one further twist. The products are divided into two classes of equal size according to a rather complicated rule and then the sum is formed with the terms in one class multiplied by $+1$ and those in the other class multiplied by -1.

Here is the definition again for $n = 3$ arranged to exhibit the signs.

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \tag{11}$$
$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$- a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

The definition for $n = 4$ involves $4! = 24$ terms, and I won't bother to write it out.

A better way to develop the theory is *recursively.* That is, we assume that determinants have been defined for all $(n-1) \times (n-1)$ matrices, and then use this to define determinants for $n \times n$ matrices. Since we have a definition for 1×1 matrices, this allows us in principle to find the determinant of any $n \times n$ matrix by recursively invoking the definition. This is less explicit, but it is easier to work with.

Here is the recursive definition. Let A be an $n \times n$ matrix, and let $D_j(A)$ be the determinant of the $(n-1) \times (n-1)$ matrix obtained by *deleting* the jth *row* and the *first column* of A. Then, define

$$\det A = a_{11}D_1(A) - a_{21}D_2(A) + \cdots + (-1)^{j+1}a_{j1}D_j(A) + \cdots + (-1)^{n+1}a_{n1}D_n(A).$$

In words: take each entry in the first column of A, multiply it by the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the first column and that row, and then add up these entries alternating signs as you do.

Examples.

$$\det \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 0 \\ 0 & 3 & 6 \end{bmatrix} = 2 \det \begin{bmatrix} 2 & 0 \\ 3 & 6 \end{bmatrix} - 1 \det \begin{bmatrix} -1 & 3 \\ 3 & 6 \end{bmatrix} + 0 \det \begin{bmatrix} -1 & 3 \\ 2 & 0 \end{bmatrix}$$
$$= 2(12 - 0) - 1(-6 - 9) + 0(\dots) = 24 + 15 = 39.$$

Note that we didn't bother evaluating the 2×2 determinant with coefficient 0. You should check that the earlier definition gives the same result.

$$
\begin{aligned}
(12) \quad \det \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & 2 & 0 \\ 2 & 0 & 3 & 6 \\ 1 & 1 & 2 & 1 \end{bmatrix}
&= 1 \det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 6 \\ 1 & 2 & 1 \end{bmatrix} - 0 \det \begin{bmatrix} 2 & -1 & 3 \\ 0 & 3 & 6 \\ 1 & 2 & 1 \end{bmatrix} \\
&\quad + 2 \det \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \end{bmatrix} - 1 \det \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 0 \\ 0 & 3 & 6 \end{bmatrix}.
\end{aligned}
$$

Each of these 3×3 determinants may be evaluated recursively. (In fact we just did the last one in the previous example.) You should work them out for yourself. The answers yield

$$
\det \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & 2 & 0 \\ 2 & 0 & 3 & 6 \\ 1 & 1 & 2 & 1 \end{bmatrix} = 1(3) - 0(\dots) + 2(5) - 1(39) = -26.
$$

Although this definition allows us to compute the determinant of any $n \times n$ matrix *in principle*, the number of operations grows very quickly with n. In such calculations one usually keeps track only of the multiplications since they are usually the most time consuming operations. Here are some values of $N(n)$, the number of multiplications needed for a recursive calculation of the determinant of an $n \times n$ determinant. We also tabulate $n!$ for comparison.

n	$N(n)$	$n!$
2	2	2
3	6	6
4	28	24
5	145	120
6	876	720
$\vdots$	$\vdots$	$\vdots$

The recursive method is somewhat more efficient than the formula referred to at the beginning of the section. For, that formula has $n!$ terms, each of which requires multiplying n entries together. Each such product requires $n-1$ separate multiplications. Hence, there are $(n-1)n!$ multiplications required altogether. In addition, the rule for determining the *sign* of the term requires some extensive calculation. However, as the above table indicates, even the number $N(n)$ of multiplications required for the recursive definition grows faster than $n!$, so it gets very large very quickly. Thus, we clearly need a more efficient method to calculate determinants. As is often the case in linear algebra, elementary row operations come to our rescue.

Using row operations for calculating determinants is based on the following rules relating such operations to determinants.

Rule (i): *If A' is obtained from A by adding a multiple of one row of A to another, then* $\det A' = \det A$.

EXAMPLE 2.1.

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & 2 & 1 \end{bmatrix} = 1(1-6) - 2(2-6) + 1(6-3) = 6$$
$$\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 1 & 2 & 1 \end{bmatrix} = 1(-3+6) - 0(2-6) + 1(-6+9) = 6.$$

Rule (ii): *if A' is obtained from A by multiplying one row by a scalar c, then* $\det A' = c \det A$.

EXAMPLE 2.2.

$$\det \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 1 & 1 \end{bmatrix} = 1(4-2) - 2(2-0) + 0(\dots) = -2$$
$$2 \det \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} = 2(1(2-1) - 1(2-0) + 0(\dots)) = 2(-1) = -2.$$

One may also state this rule as follows: *any common factor of a row of A may be 'pulled out' from its determinant.*

Rule (iii): *If A' is obtained from A by interchanging two rows, then* $\det A' = -\det A$.

EXAMPLE 2.3.

$$\det \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = -3 \qquad \det \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 3.$$

The verification of these rules is a bit involved, so we relegate it to an appendix, which most of you will want to skip.

The rules allow us to compute the determinant of any $n \times n$ matrix with specific numerical entries.

EXAMPLE 2.4. We shall calculate the determinant of a 4×4 matrix. You should make sure you keep track of which elementary row operations have been performed at each stage.

$$\det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 3 & 0 & 1 & 1 \\ -1 & 6 & 0 & 2 \end{bmatrix} = \det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & -6 & 4 & -2 \\ 0 & 8 & -1 & 3 \end{bmatrix} = \det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 7 & 4 \\ 0 & 0 & -5 & -5 \end{bmatrix}$$
$$= -5 \det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 7 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix} = +5 \det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 7 & 4 \end{bmatrix}$$
$$= +5 \det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -3 \end{bmatrix}.$$

We may now use the recursive definition to calculate the last determinant. In each case there is only one non-zero entry in the first column.

$$\det \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -3 \end{bmatrix}. = 1 \det \begin{bmatrix} 2 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -3 \end{bmatrix} \tag{13}$$
$$= 1 \cdot 2 \det \begin{bmatrix} 1 & 1 \\ 0 & -3 \end{bmatrix} = 1 \cdot 2 \cdot 1 \det \begin{bmatrix} -3 \end{bmatrix}$$
$$= 1 \cdot 2 \cdot 1 \cdot (-3) = -6.$$

Hence, the determinant of the original matrix is $5(-6) = -30$.

The last calculation is a special case of a general fact which is established in much the same way by repeating the recursive definition.

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} = a_{11}a_{22}a_{33} \dots a_{nn}.$$

In words, *the determinant of an upper triangular matrix is the product of its diagonal entries.*

It is important to be able to tell when the determinant of an $n \times n$ matrix A is zero. Certainly, this will be the case if the first column consists of zeroes, and indeed it turns out that the determinant vanishes if any row or any column consists only of zeroes. More generally, if either the set of rows or the set of columns is a linearly *dependent* set, then the determinant is zero. (That will be the case if the rank $r < n$ since the rank is the dimension of both the row space and the column space.) This follows from the following important theorem.

THEOREM 2.1. *Let A be an $n \times n$ matrix. Then A is singular if and only if* $\det A = 0$. *Equivalently, A is invertible, i.e., has rank n, if and only if* $\det A \neq 0$.

PROOF. If A is invertible, then Gaussian reduction leads to an upper triangular matrix with non-zero entries on its diagonal, and the determinant of such a matrix is the product of its diagonal entries, which is also non-zero. No elementary row operation can make the determinant zero. For, type (i) operations don't change the determinant, type (ii) operations multiply by non-zero scalars, and type (iii) operations change its sign. Hence, $\det A \neq 0$.

If A is singular, then Gaussian reduction also leads to an upper triangular matrix, but one in which at least the last row consists of zeroes. Hence, at least one diagonal entry is zero, and so is the determinant. $\square$

Example 2.5.

$$\det \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} = 1(1-0) - 2(1-0) + 1(3-2) = 0$$

so the matrix must be singular. To confirm this, we reduce

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

which shows that the matrix is singular.

In the previous section, we encountered 2×2 matrices with symbolic non-numeric entries. For such a matrix, Gaussian reduction doesn't work very well because we don't know whether the non-numeric expressions are zero or not.

Example 2.6. Suppose we want to know whether or not the matrix

$$\begin{bmatrix} -\lambda & 1 & 1 & 1 \\ 1 & -\lambda & 0 & 0 \\ 1 & 0 & -\lambda & 0 \\ 1 & 0 & 0 & -\lambda \end{bmatrix}$$

is singular. We could try to calculate its rank, but since we don't know what λ is, it is not clear how to proceed. Clearly, the row reduction works differently if $\lambda = 0$ than if $\lambda \neq 0$. However, we can calculate the determinant by the recursive method.

$$\begin{aligned}
(14) \quad & \det \begin{bmatrix} -\lambda & 1 & 1 & 1 \\ 1 & -\lambda & 0 & 0 \\ 1 & 0 & -\lambda & 0 \\ 1 & 0 & 0 & -\lambda \end{bmatrix} \\
&= (-\lambda) \det \begin{bmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{bmatrix} - 1 \det \begin{bmatrix} 1 & 1 & 1 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{bmatrix} \\
&\quad + 1 \det \begin{bmatrix} 1 & 1 & 1 \\ -\lambda & 0 & 0 \\ 0 & 0 & -\lambda \end{bmatrix} - 1 \det \begin{bmatrix} 1 & 1 & 1 \\ -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \end{bmatrix} \\
&= (-\lambda)(-\lambda^3) - (\lambda^2) + (-\lambda^2) - (\lambda^2) \\
&= \lambda^4 - 3\lambda^2 = \lambda^2(\lambda - \sqrt{3})(\lambda + \sqrt{3}).
\end{aligned}$$

Hence, this matrix is singular just in the cases $\lambda = 0, \lambda = \sqrt{3}$, and $\lambda = -\sqrt{3}$.

Appendix. Some Proofs. We now establish the basic rules relating determinants to elementary row operations. If you are of a skeptical turn of mind, you should study this section, since the relation between the recursive definition and rules (i), (ii), and (iii) is not at all obvious. However, if you have a trusting nature, you might want to skip this section since the proofs are quite technical and not terribly enlightening.

The idea behind the proofs is to assume that the rules—actually, modified forms of the rules—have been established for $(n-1)\times(n-1)$ determinants, and then to prove them for $n\times n$ determinants. To start it all off, the rules must be checked explicitly for 2×2 determinants. I leave that step for you in the Exercises.

We start with the hardest case, rule (iii). First we consider the special case that A' is obtained from A by switching two *adjacent* rows, the ith row and the $(i+1)$st row. Consider the recursive definition

$$\begin{aligned}(15)\quad \det A' = a'_{11}D_1(A') - \cdots + (-1)^{i+1}a'_{i1}D_i(A')\\ + (-1)^{i+2}a'_{i+1,1}D_{i+1}(A') + \cdots + (-1)^{n+1}a'_{n1}D_n(A').\end{aligned}$$

Look at the subdeterminants occurring in this sum. For $j \neq i, i+1$, we have

$$D_j(A') = -D_j(A)$$

since deleting the first column and jth row of A and then switching two rows—neither of which was deleted—changes the sign by rule (iii) for $(n-1)\times(n-1)$ determinants. The situation for $j = i$ or $j = i+1$ is different; in fact, we have

$$D_i(A') = D_{i+1}(A) \qquad \text{and} \qquad D_{i+1}(A') = D_i(A).$$

The first equation follows because switching rows i and $i+1$ and then deleting row i is the same as deleting row $i+1$ without touching row i. A similar argument establishes the second equation. Using this together with $a'_{i1} = a_{i+1,1}, a'_{i+1,1} = a_{i1}$ yields

$$\begin{aligned}(-1)^{i+1}a'_{i1}D_i(A') &= (-1)^{i+1}a_{i+1,1}D_{i+1}(A) = -(-1)^{i+2}a_{i+1,1}D_{i+1}(A)\\ (-1)^{i+2}a'_{i+1,1}D_{i+1}(A') &= (-1)^{i+2}a_{i1}D_i(A) = -(-1)^{i+1}a_{i1}D_i(A).\end{aligned}$$

In other words, all terms in the recursive definition of $\det A'$ are negatives of the corresponding terms of $\det A$ *except* those in positions i and $i+1$ which get reversed with signs changed. Hence, the effect of switching adjacent rows is to change the sign of the sum.

Suppose instead that non-adjacent rows in positions i and j are switched, and suppose for the sake of argument that $i < j$. One way to do this is as follows. First move row i past each of the rows between row i and row j. This involves some number of switches of adjacent rows—call that number k. ($k = j - i - 1$, but it that doesn't matter in the proof.) Next, move row j past row i and then past the k rows just mentioned, all in their new positions. That requires $k+1$ switches of adjacent rows. All told, to switch rows i and j in this way requires $2k+1$ switches of adjacent rows. The net effect is to multiply the determinant by $(-1)^{2k+1} = -1$ as required.

There is one important consequence of rule (iii) which we shall use later in the proof of rule (i).

Rule (iiie): *If an $n\times n$ matrix has two equal rows, then* $\det A = 0$.

This is not too hard to see. Interchanging two rows changes the sign of $\det A$, but if the rows are equal, it doesn't change anything. However, the only number with the property that it isn't changed by changing its sign is the number 0. Hence, $\det A = 0$. We next verify rule (ii). Suppose A' is obtained from A by multiplying the ith row by c. Consider the recursive definition

$$\det A' = a'_{11}D_1(A') + \cdots + (-1)^{i+1}a'_{i1}D_i(A') + \cdots + (-1)^{n+1}a_{n1}D_n(A). \tag{16}$$

For any $j \neq i$, $D_j(A') = cD_j(A)$ since one of the rows appearing in that determinant is multiplied by c. Also, $a'_{j1} = a_{j1}$ for $j \neq i$. On the other hand, $D_i(A') = D_i(A)$ since the ith row is deleted in calculating these quantities, and, except for the ith row, A' and A agree. In addition, $a'_{i1} = ca_{i1}$ so we pick up the extra factor of c in any case. It follows that every term on the right of (16) has a factor c, so $\det A' = c \det A$.

Finally, we attack the proof of rule (i). It turns out to be necessary to verify the following stronger rule.

Rule (ia): *Suppose A, A', and A'' are three $n \times n$ matrices which agree except in the ith row. Suppose moreover that the ith row of A is the sum of the ith row of A' and the ith row of A''. Then* $\det A = \det A' + \det A''$.

Let's first see why rule (ia) implies rule (i). We can add c times the jth row of A to its i row as follows. Let $B' = A$, let B'' be the matrix obtained from A by replacing its ith row by c times its jth row, and let B be the matrix obtained form A by adding c times its jth row to its ith row. Then according to rule (ia), we have

$$\det B = \det B' + \det B'' = \det A + \det B''.$$

On the other hand, by rule (ii), $\det B'' = c \det A''$ where A'' has both ith and jth rows equal to the jth row of A. Hence, by rule (iiie), $\det A'' = 0$, and $\det B = \det A$.

Finally, we establish rule (1a). Assume it is known to be true for $(n-1)\times(n-1)$ determinants. We have

$$\det A = a_{11}D_1(A) - \cdots + (-1)^{i+1}a_{i1}D_i(A) + \cdots + (-1)^{n+1}a_{n1}D_n(A). \tag{17}$$

For $j \neq i$, the the sum rule (ia) may be applied to the determinants $D_i(A)$ because the appropriate submatrix has one row which breaks up as a sum as needed. Hence,

$$D_j(A) = D_j(A') + D_j(A'').$$

Also, for $j \neq i$, we have $a_{j1} = a'_{j1} = a''_{j1}$ since all the matrices agree in any row except the ith row. Hence, for $j \neq i$,

$$a_{i1}D_i(A) = a_{i1}D_i(A') + a_{i1}D_i(A'') = a'_{i1}D_i(A') + a''_{i1}D_i(A'').$$

On the other hand, $D_i(A) = D_i(A') = D_i(A'')$ because in each case the ith row was deleted. But $a_{i1} = a'_{i1} + a''_{i1}$, so

$$a_{i1}D_i(A) = a'_{i1}D_i(A) + a''_{i1}D_i(A) = a'_{i1}D_i(A') + a''_{i1}D_i(A'').$$

It follows that every term in (17) breaks up into a sum as required, and $\det A = \det A' + \det A''$.

Exercises for Section 2.

2.1. Find the determinants of each of the following matrices. Use whatever method seems most convenient, but seriously consider the use of elementary row operations.

(a) $\begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 5 \\ 6 & 4 & 1 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 1 & 4 & 2 & 3 \\ 4 & 3 & 2 & 1 \end{bmatrix}$.

(c) $\begin{bmatrix} 0 & 0 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$.

(d) $\begin{bmatrix} 0 & x & y \\ -x & 0 & z \\ -y & -z & 0 \end{bmatrix}$.

2.2. Verify the following rules for 2×2 determinants.

(i) If A' is obtained from A by adding a multiple of the first row to the second, then $\det A' = \det A$.

(ii) If A' is obtained from A by multiplying its first row by c, then $\det A' = c \det A$.

(iii) If A' is obtained from A by interchanging its two rows, then $\det A' = -\det A$.

Rules (i) and (ii) for the first row, together with rule (iii) allow us to derive rules (i) and (ii) for the second row. Explain.

2.3. Derive the following generalization of rule (i) for 2×2 determinants.

$$\det \begin{bmatrix} a' + a'' & b' + b'' \\ c & d \end{bmatrix} = \det \begin{bmatrix} a' & b' \\ c & d \end{bmatrix} + \det \begin{bmatrix} a'' & b'' \\ c & d \end{bmatrix}.$$

What is the corresponding rule for the second row? Why do you get it for free if you use the results of the previous problem?

2.4. Find all values of z such that the matrix $\begin{bmatrix} 1 & z \\ z & 1 \end{bmatrix}$ is singular.

2.5. Find all values of λ such that the matrix

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

is singular.

2.6. The determinant of the following matrix is zero. Explain why just using the recursive definition of the determinant.

$$\begin{bmatrix} 2 & -2 & 3 & 0 & 4 \\ 4 & 2 & -3 & 0 & 1 \\ 6 & -5 & -2 & 0 & 3 \\ 1 & -3 & 3 & 0 & 6 \\ 5 & 2 & 12 & 0 & 10 \end{bmatrix}$$

2.7. If A is $n \times n$, what can you say about $\det(cA)$?

2.8. Suppose A is a non-singular 6×6 matrix. Then $\det(-A) \neq -\det A$. Explain.

2.9. Find 2×2 matrices A and B such that $\det(A + B) \neq \det A + \det B$.

2.10. (a) Show that the number of multiplications $N(7)$ necessary to compute recursively the determinant of a 7×7 matrix is 6139.

(b) (Optional) Find a rule relating $N(n)$ to $N(n-1)$. Use this to write a computer program to calculate $N(n)$ for any n.

3. Some Important Properties of Determinants

THEOREM 3.1. *(The Product Rule) Let A and B be $n \times n$ matrices. Then*

$$\det(AB) = \det A \det B.$$

We relegate the proof of this theorem to an appendix, but let's check it in an example

EXAMPLE 3.1. Let

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Then $\det A = 3, \det B = 2$, and

$$AB = \begin{bmatrix} 3 & -1 \\ 3 & 1 \end{bmatrix}$$

so $\det(AB) = 6$ as expected.

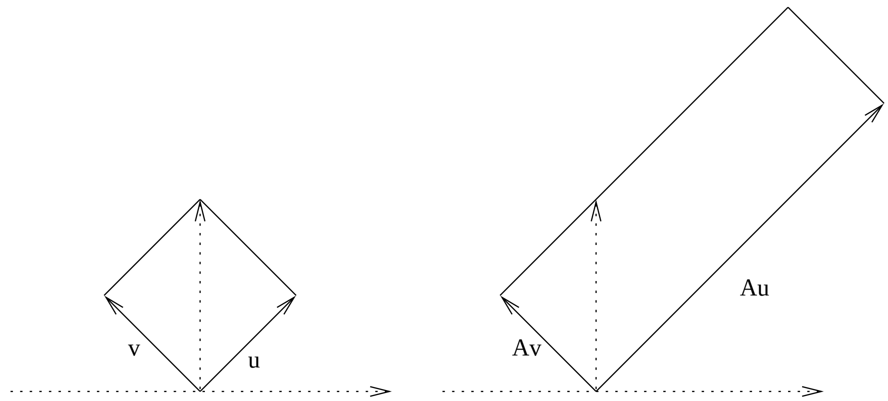
This example has a simple geometric interpretation. Let

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Then $\det B$ is just the area of the parallelogram spanned by the two vectors. On the other hand the columns of the product

$$AB = \begin{bmatrix} A\mathbf{u} & A\mathbf{v} \end{bmatrix} \qquad \text{i.e., } A\mathbf{u} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad A\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

also span a parallelogram which is related to the first parallelogram in a simple way. One edge is multiplied by a factor 3 and the other edge is fixed. Hence, the area is multiplied by 3.



Thus, in this case, in the formula

$$\det(AB) = (\det A)(\det B)$$

the factor $\det A$ tells us how the area of a parallelogram changes if its edges are transformed by the matrix A. This is a special case of a much more general assertion. The product rule tells us how areas, volumes, and their higher dimensional analogues behave when a figure is transformed by a matrix.

Transposes. Let A be an $m \times n$ matrix. The *transpose* of A is the $n \times m$ matrix for which the columns are the rows of A. (Also, its rows are the columns of A.) It is usually denoted A^t, but other notations are possible.

Examples.

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix} \qquad A^t = \begin{bmatrix} 2 & 2 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \tag{18}$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix} \qquad A^t = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 3 & 3 & 3 \end{bmatrix} \tag{19}$$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \mathbf{a}^t = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}. \tag{20}$$

Note that the transpose of a column vector is a row vector and vice-versa.

The following rule follows almost immediately from the definition.

Theorem 3.2. *Assume A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then*

$$(AB)^t = B^t A^t.$$

Note that *the order on the right is reversed.*

Example 3.2. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 0 \\ 1 & 1 \\ 2 & 3 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 12 & 11 \\ 8 & 6 \\ 4 & 0 \end{bmatrix}, \quad \text{so} \quad (AB)^t = \begin{bmatrix} 12 & 8 & 4 \\ 11 & 6 & 0 \end{bmatrix}$$

while

$$B^t = \begin{bmatrix} 4 & 1 & 2 \\ 0 & 1 & 3 \end{bmatrix}, \quad A^t = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \\ 3 & 2 & 0 \end{bmatrix}, \qquad \text{so} \quad B^tA^t = \begin{bmatrix} 12 & 8 & 4 \\ 11 & 6 & 0 \end{bmatrix}$$

as expected.

Unless the matrices are square, the shapes won't even match if the order is not reversed. In the above example A^tB^t would be a product of a 3×3 matrix with a 2×3 matrix, and that doesn't make sense. The example also helps us to understand why the formula is true. The i, j-entry of the product is the row by column product of the ith row of A with the jth column of B. However, taking transposes reverses the roles of rows and columns. The entry is the same, but now it is the product of the jth row of B^t with the ith column of A^t.

Theorem 3.3. *Let A be an $n \times n$ matrix. Then*

$$\det A^t = \det A.$$

See the appendix for a proof, but here is an example.

Example 3.3.

$$\det \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = 1(1-0) - 2(0-0) + 0(\dots) = 1$$

$$\det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} = 1(1-0) - 0(\dots) + 1(0-0) = 1.$$

The importance of this theorem is that it allows us to go freely from statements about determinants involving rows of the matrix to corresponding statements involving columns and vice-versa.

Because of this rule, we may use *column operations* as well as row operations to calculate determinants. For, performing a column operation is the same as transposing the matrix, performing the corresponding row operation, and then transposing back. The two transpositions don't affect the determinant.

Example 3.4. Example

$$\det \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 3 & 1 \\ 3 & 3 & 6 & 2 \\ 4 & 2 & 6 & 4 \end{bmatrix} = \det \begin{bmatrix} 1 & 2 & 2 & 0 \\ 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 2 \\ 4 & 2 & 2 & 4 \end{bmatrix} \qquad \text{operation } (-1)c1 + \mathbf{c3}$$

$$= 0.$$

The last step follows because the 2nd and 3rd columns are equal, which implies that the rank (dimension of the column space) is less than 4. (You could also subtract the third column from the second and get a column of zeroes, etc.)

Expansion in Minors or Cofactors. There is a generalization of the formula used for the recursive definition. Namely, for any $n \times n$ matrix A, let $D_{ij}(A)$ be the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A. Then,

(21)

$$\begin{aligned} \det A &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} D_{ij}(A) \\ &= (-1)^{1+j} a_{1j} D_{1j}(A) + \cdots + (-1)^{i+j} a_{ij} D_{ij}(A) + \cdots + (-1)^{n+j} a_{nj} D_{nj}(A). \end{aligned}$$

The special case $j = 1$ is the recursive definition given in the previous section. The more general rule is easy to derive from the special case $j = 1$ by means of column interchanges. Namely, form a new matrix A' by moving the jth column to the first position by successively interchanging it with columns $j-1, j-2, \dots, 2, 1$. There are $j-1$ interchanges, so the determinant is changed by the factor $(-1)^{j-1}$. Now apply the rule for the first column. The first column of A' is the jth column of A, and deleting it has the same effect as deleting the jth column of A. Hence, $a'_{i1} = a_{ij}$ and $D_i(A') = D_{ij}(A)$. Thus,

$$\begin{aligned} \det A = (-1)^{j-1} \det A' &= (-1)^{j-1} \sum_{i=1}^{n} (-1)^{1+i} a'_{i1} D_i(A') \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} D_{ij}(A). \end{aligned}$$

Similarly, there is a corresponding rule for any *row* of a matrix

$$\begin{aligned} \text{(22)}\quad \det A &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} D_{ij}(A) \\ &= (-1)^{i+1} a_{i1} D_{i1} + \cdots + (-1)^{i+j} a_{ij} D_{ij}(A) + \cdots + (-1)^{i+n} a_{in} D_{in}(A). \end{aligned}$$

This formula is obtained from (21) by transposing, applying the corresponding column rule, and then transposing back.

Example 3.5. Expand the following determinant using its second row.

$$\text{(23)}\quad \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 6 & 0 \\ 3 & 2 & 1 \end{bmatrix} = (-1)^{2+3} 0(\dots) + (-1)^{2+2} 6 \det \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} + (-1)^{2+3} 0(\dots)$$
$$= 6(1-9) = -48.$$

There is some terminology which you may see used in connection with these formulas. The determinant $D_{ij}(A)$ of the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column is called the *i, j-minor* of A. The quantity $(-1)^{i+j} D_{ij}(A)$ is called the *i, j-cofactor*. Formula (21) is called expansion in minors (or cofactors) of the jth column and formula (22) is called expansion in minors (or cofactors) of the ith row. It is not necessary to remember the terminology as long as you remember the formulas and understand how they are used.

Cramer's Rule. One may use determinants to derive a formula for the solutions of a *non-singular* system of n equations in n unknowns

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The formula is called *Cramer's rule*, and here it is. *For the jth unknown x_j, take the determinant of the matrix formed by replacing the jth column of the coefficient matrix A by* $\mathbf{b}$*, and divide it by* $\det A$. In symbols,

$$x_j = \frac{\det \begin{bmatrix} a_{11} & \dots & b_1 & \dots & a_{1n} \\ a_{21} & \dots & b_2 & \dots & a_{2n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ a_{n1} & \dots & b_n & \dots & a_{nn} \end{bmatrix}}{\det \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix}}$$

EXAMPLE 3.6. Consider

$$\begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 2 \\ 2 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix}.$$

We have

$$\det \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 2 \\ 2 & 0 & 6 \end{bmatrix} = 2.$$

(Do you see a quick way to compute that?) Hence,

$$x_1 = \frac{\det \begin{bmatrix} 1 & 0 & 2 \\ 5 & 1 & 2 \\ 3 & 0 & 6 \end{bmatrix}}{2} = \frac{0}{2} = 0$$

$$x_2 = \frac{\det \begin{bmatrix} 1 & 1 & 2 \\ 1 & 5 & 2 \\ 2 & 3 & 6 \end{bmatrix}}{2} = \frac{8}{2} = 4$$

$$x_3 = \frac{\det \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 5 \\ 2 & 0 & 3 \end{bmatrix}}{2} = \frac{1}{2}.$$

You should try to do this by Gauss-Jordan reduction.

Cramer's rule is not too useful for solving specific numerical systems of equations. The only practical method for calculating the needed determinants for n large is to use row (and possibly column) operations. It is usually easier to use row operations to solve the system without resorting to determinants. However, if the system has non-numeric symbolic coefficients, Cramer's rule is sometimes useful. Also, it is often valuable as a theoretical tool.

Cramer's rule is related to expansion in minors. You can find further discussion of it and proofs in Section 5.4 and 5.5 of *Introduction to Linear Algebra* by Johnson, Riess, and Arnold. (See also Section 4.5 of *Applied Linear Algebra* by Noble and Daniel.)

Appendix. Some Proofs. Here are the proofs of two important theorems stated in this section.

The Product Rule. $\det(AB) = (\det A)(\det B)$.

Proof. First assume that A is non-singular. Then there is a sequence of row operations which reduces A to the identity

$$A \to A_1 \to A_2 \to \ldots \to A_k = I.$$

Associated with each of these operations will be a multiplier c_i which will depend on the particular operation, and

$$\det A = c_1 \det A_1 = c_1 c_2 \det A_2 = \cdots = c_1 c_2 \ldots c_k \det A_k = c_1 c_2 \ldots c_k$$

since $A_k = I$ and $\det I = 1$. Now apply exactly these row operations to the product AB

$$AB \to A_1 B \to A_2 B \to \ldots \to A_k B = IB = B.$$

The same multipliers contribute factors at each stage, and

$$\det AB = c_1 \det A_1 B = c_1 c_2 \det A_2 B = \cdots = \underbrace{c_1 c_2 \ldots c_k}_{\det A} \det B = \det A \ \det B.$$

Assume instead that A is singular. Then, AB is also singular. (This follows from the fact that the rank of AB is at most the rank of A, as mentioned in the Exercises for Chapter 1, Section 6. However, here is a direct proof for the record. Choose a sequence of elementary row operations for A, the end result of which is a matrix A' with at least one row of zeroes. Applying the same operations to AB yields $A'B$ which also has to have at least one row of zeroes.) It follows that both $\det AB$ and $\det A \det B$ are zero, so they are equal. $\square$

The Transpose Rule. $\det A^t = \det A$.

Proof. If A is singular, then A^t is also singular and vice-versa. For, the rank may be characterized as either the dimension of the row space or the dimension of the column space, and an $n \times n$ matrix is singular if its rank is less than n. Hence, in the singular case, $\det A = 0 = \det A^t$.

Suppose then that A is non-singular. Then there is a sequence of elementary row operations

$$A \to A_1 \to A_2 \to \cdots \to A_k = I.$$

Recall from Chapter 1, Section 4 that each elementary row operation may be accomplished by multiplying by an appropriate elementary matrix. Let C_i denote the elementary matrix needed to perform the ith row operation. Then,

$$A \to A_1 = C_1 A \to A_2 = C_2 C_1 A \to \cdots \to A_k = C_k C_{k-1} \ldots C_2 C_1 A = I.$$

In other words,

$$A = (C_k \ldots C_2 C_1)^{-1} = {C_1}^{-1} {C_2}^{-1} \ldots {C_k}^{-1}.$$

To simplify the notation, let $D_i = {C_i}^{-1}$. The inverse D of an elementary matrix C is also an elementary matrix; its effect is the row operation which reverses the effect of C. Hence, we have shown that *any non-singular square matrix A may be expressed as a product of elementary matrices*

$$A = D_1 D_2 \dots D_k.$$

Hence, by the product rule

$$\det A = (\det D_1)(\det D_2)\dots(\det D_k).$$

On the other hand, we have by rule for the transpose of a product

$$A^t = {D_k}^t \dots {D_2}^t {D_1}^t,$$

so by the product rule

$$\det A^t = \det({D_k}^t)\dots\det({D_2}^t)\det({D_1}^t).$$

Suppose we know the rule $\det D^t = \det D$ *for any elementary matrix D*. Then,

$$\begin{aligned}\det A^t &= \det({D_k}^t)\dots\det({D_2}^t)\det({D_1}^t)\\ &= \det(D_k)\dots\det(D_2)\det(D_1)\\ &= (\det D_1)(\det D_2)\dots(\det D_k) = \det A.\end{aligned}$$

(We used the fact that the products on the right are products of scalars and so can be rearranged any way we like.)

It remains to establish the rule for elementary matrices. If $D = E_{ij}(c)$ is obtained from the identity matrix by adding c times its jth row to its ith row, then $D^t = E_{ji}(c)$ is a matrix of exactly the same type. In each case, $\det D = \det D^t = 1$. If $D = E_i(c)$ is obtained by multiplying the ith row of the identity matrix by c, then D^t is exactly the same matrix $E_i(c)$. Finally, if $D = E_{ij}$ is obtained from the identity matrix by interchanging its ith and jth rows, then D^t is E_{ji} which in fact is just E_{ij} again. Hence, in each case $\det D^t = \det D$ does hold. □

Exercises for Section 3.

3.1. Check the validity of the product rule for the product

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

3.2. If A and B are $n \times n$ matrices, both of rank n, what can you say about the rank of AB?

3.3. Find

$$\det \begin{bmatrix} 3 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 6 & 4 & 0 \\ 1 & 5 & 4 & 3 \end{bmatrix}.$$

Of course, the answer is the product of the diagonal entries. Using the properties discussed in the section, see how many different ways you can come to this conclusion.

What can you conclude in general about the determinant of a lower triangular square matrix?

3.4. (a) Show that if A is an invertible $n \times n$ matrix, then $\det(A^{-1}) = \dfrac{1}{\det A}$. Hint: Let $B = A^{-1}$ and apply the product rule to AB.

(b) Using part(a), show that if A is any $n \times n$ matrix and P is an invertible $n \times n$ matrix, then $\det(PAP^{-1}) = \det A$.

3.5. Why does Cramer's rule fail if the coefficient matrix A is singular?

3.6. Use Cramer's rule to solve the system

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

Also, solve it by Gauss-Jordan reduction and compare the amount of work you had to do in each case.

4. Eigenvalues and Eigenvectors

One way in which to understand a matrix A is to examine its effects on the geometry of vectors in $\mathbf{R}^n$. For example, we saw that $\det A$ measures the relative change in area or volume for figures generated by vectors in two or three dimensions. Also, as we have seen in an exercise in Chapter 1, Section 2, the multiples $A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_n$ of the standard basis vectors are just the columns of the matrix A. More generally, it is often useful to look at multiples $A\mathbf{v}$ for other vectors $\mathbf{v}$.

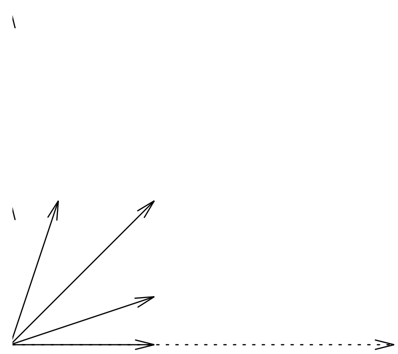
Example 4.1. Let

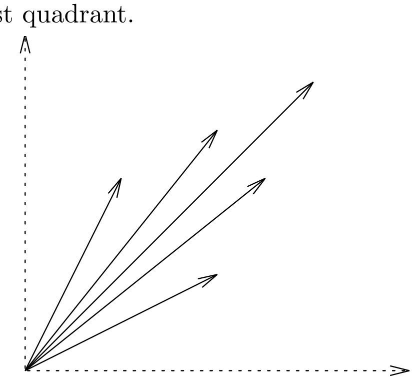
$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Here are some examples of products $A\mathbf{v}$.

$\mathbf{v}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$A\mathbf{v}$	$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 2.5 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 2.5 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

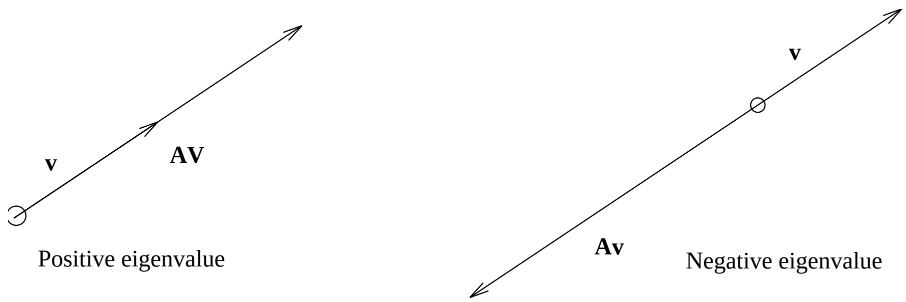
These illustrate a trend for vectors in the first quadrant.





Vectors pointing near one or the other of the two axes are directed closer to the diagonal line. A diagonal vector is transformed into another diagonal vector.

Let A be any $n \times n$ matrix. In general, if $\mathbf{v}$ is a vector in $\mathbf{R}^n$, the transformed vector $A\mathbf{v}$ will differ from $\mathbf{v}$ in both magnitude and direction. However, some vectors $\mathbf{v}$ will have the property that $A\mathbf{v}$ ends up being *parallel* to $\mathbf{v}$; i.e., it points in the same direction or the opposite direction. These vectors will specify 'natural' axes for any problem involving the matrix A.



Vectors are parallel when one is a scalar multiple of the other, so we make the following formal definition. A *non-zero* vector $\mathbf{v}$ is called an *eigenvector* for the square matrix A if

$$A\mathbf{v} = \lambda\mathbf{v} \tag{24}$$

for an appropriate scalar λ. λ is called the *eigenvalue* associated with the eigenvector $\mathbf{v}$.

In words, this says that $\mathbf{v} \neq \mathbf{0}$ is an eigenvector for A if multiplying it by A has the same effect as multiplying it by an appropriate scalar. Thus, we may think of eigenvectors as being vectors for which matrix multiplication by A takes on a particularly simple form.

It is important to note that while the eigenvector $\mathbf{v}$ must be *non-zero*, the corresponding eigenvalue λ is allowed to be zero.

Example 4.2. Let

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

We want to see if the system

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

has non-trivial solutions v_1, v_2. This of course depends on λ. If we write this system out, it becomes

$$\begin{aligned} 2v_1 + v_2 &= \lambda v_1 \\ v_1 + 2v_2 &= \lambda v_2 \end{aligned}$$

or, collecting terms,

$$\begin{aligned} (2-\lambda)v_1 + v_2 &= 0 \\ v_1 + (2-\lambda)v_2 &= 0. \end{aligned}$$

In matrix form, this becomes

$$\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \mathbf{v} = \mathbf{0}. \tag{25}$$

For any specific λ, this is a homogeneous system of two equations in two unknowns. By the theory developed in the previous sections, we know that it will have non-zero solutions precisely in the case that the rank is smaller than two. A simple criterion for that to be the case is that the determinant of the coefficient matrix should vanish, i.e.,

$$\det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = (2-\lambda)^2 - 1 = 0$$
$$\text{or} \qquad 4 - 4\lambda + \lambda^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1) = 0.$$

The roots of this equation are $\lambda = 3$ and $\lambda = 1$. Thus, these and only these scalars λ can be eigenvalues for appropriate eigenvectors.

First consider $\lambda = 3$. Putting this in (25) yields

$$\begin{bmatrix} 2-3 & 1 \\ 1 & 2-3 \end{bmatrix} \mathbf{v} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{v} = \mathbf{0}. \tag{26}$$

Gauss-Jordan reduction yields

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

(As is usual for homogeneous systems, we don't need to explicitly write down the augmented matrix, because there are zeroes to the right of the 'bar'.) The corresponding system is $v_1 - v_2 = 0$, and the general solution is $v_1 = v_2$ with v_2 free. A general solution vector has the form

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_2 \\ v_2 \end{bmatrix} = v_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Put $v_2 = 1$ to obtain

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which will form a basis for the solution space of (26). Any other eigenvector for $\lambda = 3$ will be a non-zero multiple of the basis vector $\mathbf{v}_1$.

Consider next the eigenvalue $\lambda = 1$. Put this in (25) to obtain

$$\begin{bmatrix} 2-1 & 1 \\ 1 & 2-1 \end{bmatrix} \mathbf{v} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{v} = \mathbf{0}. \tag{27}$$

In this case, Gauss-Jordan reduction—which we omit—yields the general solution $v_1 = -v_2$ with v_2 free. The general solution vector is

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Putting $v_2 = 1$ yields the basic eigenvector

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The general case

We redo the above algebra for an arbitrary $n \times n$ matrix. First, rewrite the eigenvector condition as follows

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ A\mathbf{v} - \lambda\mathbf{v} &= 0 \\ A\mathbf{v} - \lambda I\mathbf{v} &= 0 \\ (A - \lambda I)\mathbf{v} &= 0. \end{aligned}$$

The last equation is the homogeneous $n \times n$ system with $n \times n$ coefficient matrix

$$A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{bmatrix}.$$

It has a *non-zero* solution vector $\mathbf{v}$ if and only if the coefficient matrix has rank less than n, i.e., if and only if it is *singular*. By Theorem 2.1, this will be true if and only if λ satisfies the *characteristic equation*

$$\det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{bmatrix} = 0. \tag{28}$$

As in the example, the strategy for finding eigenvalues and eigenvectors is as follows. First find the roots of the characteristic equation. These are the eigenvalues. Then for each root λ, find a general solution for the system

$$(A - \lambda I)\mathbf{v} = 0. \tag{29}$$

This gives us all the eigenvectors for that eigenvalue. The solution space of the system (29), i.e., the null space of the matrix $A - \lambda I$, is called the *eigenspace* corresponding to the eigenvalue λ.

EXAMPLE 4.3. Consider the matrix

$$A = \begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

The characteristic equation is

$$\begin{aligned} \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 4 & 3 \\ 4 & 1-\lambda & 0 \\ 3 & 0 & 1-\lambda \end{bmatrix} & \\ &= (1-\lambda)((1-\lambda)^2 - 0) - 4(4(1-\lambda) - 0) + 3(0 - 3(1-\lambda) \\ &= (1-\lambda)^3 - 25(1-\lambda) = (1-\lambda)((1-\lambda)^2 - 25) \\ &= (1-\lambda)(\lambda^2 - 2\lambda - 24) = (1-\lambda)(\lambda - 6)(\lambda + 4) = 0. \end{aligned}$$

Hence, the eigenvalues are $\lambda = 1, \lambda = 6$, and $\lambda = -4$. We proceed to find the eigenspaces for each of these eigenvalues, starting with the largest.

First, take $\lambda = 6$, and put it in (29) to obtain the system

$$\begin{bmatrix} 1-6 & 4 & 3 \\ 4 & 1-6 & 0 \\ 3 & 0 & 1-6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \qquad \text{or} \qquad \begin{bmatrix} -5 & 4 & 3 \\ 4 & -5 & 0 \\ 3 & 0 & -5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0.$$

To solve, use Gauss-Jordan reduction

$$\begin{aligned} \begin{bmatrix} -5 & 4 & 3 \\ 4 & -5 & 0 \\ 3 & 0 & -5 \end{bmatrix} &\to \begin{bmatrix} -1 & -1 & 3 \\ 4 & -5 & 0 \\ 3 & 0 & -5 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & 3 \\ 0 & -9 & 12 \\ 0 & -3 & 4 \end{bmatrix} \\ &\to \begin{bmatrix} -1 & -1 & 3 \\ 0 & 0 & 0 \\ 0 & -3 & 4 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & 3 \\ 0 & 3 & -4 \\ 0 & 0 & 0 \end{bmatrix} \\ &\to \begin{bmatrix} 1 & 1 & -3 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -5/3 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Note that the matrix is singular, and the rank is smaller than 3. This must be the case because the condition $\det(A - \lambda I) = 0$ guarantees it. If the coefficient matrix were non-singular, you would know that there was a mistake: either the roots of the characteristic equation are wrong or the row reduction was not done correctly.

The general solution is

$$\begin{aligned} v_1 &= (5/3)v_3 \\ v_2 &= (4/3)v_3 \end{aligned}$$

with v_3 free. The general solution vector is

$$\mathbf{v} = \begin{bmatrix} (5/3)v_3 \\ (4/3)v_3 \\ v_3 \end{bmatrix} = v_3 \begin{bmatrix} 5/3 \\ 4/3 \\ 1 \end{bmatrix}.$$

Hence, the eigenspace is 1-dimensional. A basis may be obtained by setting $v_3 = 1$ as usual, but it is a bit neater to put $v_3 = 3$ so as to avoid fractions. Thus,

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}$$

constitutes a basis for the eigenspace corresponding to the eigenvalue $\lambda = 6$. Note that we have now found all eigenvectors for this eigenvalue. They are all the non-zero vectors in this 1-dimensional eigenspace, i.e., all non-zero multiples of $\mathbf{v}_1$.

Next take $\lambda = 1$ and put it in (29) to obtain the system

$$\begin{bmatrix} 0 & 4 & 3 \\ 4 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0.$$

Use Gauss-Jordan reduction

$$\begin{bmatrix} 0 & 4 & 3 \\ 4 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3/4 \\ 0 & 0 & 0 \end{bmatrix}.$$

The general solution is

$$\begin{aligned} v_1 &= 0 \\ v_2 &= -(3/4)v_3 \end{aligned}$$

with v_2 free. Thus the general solution vector is

$$\mathbf{v} = \begin{bmatrix} 0 \\ -(3/4)v_3 \\ v_3 \end{bmatrix} = v_3 \begin{bmatrix} 0 \\ -3/4 \\ 1 \end{bmatrix}.$$

Put $v_3 = 4$ to obtain a single basis vector

$$\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 4 \end{bmatrix}$$

for the eigenspace corresponding to the eigenvalue $\lambda = 1$. The set of eigenvectors for this eigenvalue is the set of non-zero multiples of $\mathbf{v}_2$.

Finally, take $\lambda = -4$, and put this in (29) to obtain the system

$$\begin{bmatrix} 5 & 4 & 3 \\ 4 & 5 & 0 \\ 3 & 0 & 5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0.$$

Solve this by Gauss-Jordan reduction.

$$\begin{aligned} \begin{bmatrix} 5 & 4 & 3 \\ 4 & 5 & 0 \\ 3 & 0 & 5 \end{bmatrix} &\to \begin{bmatrix} 1 & -1 & 3 \\ 4 & 5 & 0 \\ 3 & 0 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 3 \\ 0 & 9 & -12 \\ 0 & 3 & -4 \end{bmatrix} \\ &\to \begin{bmatrix} 1 & -1 & 3 \\ 0 & 3 & -4 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 5/3 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The general solution is

$$\begin{aligned} v_1 &= -(5/3)v_3 \\ v_2 &= (4/3)v_3 \end{aligned}$$

with v_3 free. The general solution vector is

$$\mathbf{v} = \begin{bmatrix} -(5/3)v_3 \\ (4/3)v_3 \\ v_3 \end{bmatrix} = v_3 \begin{bmatrix} -5/3 \\ 4/3 \\ 1 \end{bmatrix}.$$

Setting $v_3 = 3$ yields the basis vector

$$\mathbf{v}_3 = \begin{bmatrix} -5 \\ 4 \\ 3 \end{bmatrix}$$

for the eigenspace corresponding to $\lambda = -4$. The set of eigenvectors for this eigenvalue consists of all non-zero multiples of $\mathbf{v}_3$.

The set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ obtained in the previous example is linearly independent. To see this apply Gaussian reduction to the matrix with these vectors as columns:

$$\begin{bmatrix} 5 & 0 & -5 \\ 4 & -3 & 4 \\ 3 & 4 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & -3 & 8 \\ 0 & 4 & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -8/3 \\ 0 & 0 & 50/3 \end{bmatrix}.$$

The reduced matrix has rank 3, so the columns of the original matrix form an independent set.

It is no accident that a set so obtained is linearly independent. The following theorem tells us that this will always be the case.

THEOREM 4.1. *Let A be an $n \times n$ matrix. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be different eigenvalues of A, and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ be corresponding eigenvectors. Then*

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$$

is a linearly independent set.

See the appendix if you are interested in a proof.

Historical Aside. The concepts discussed here were invented by the 19th century English mathematicians Cayley and Sylvester, but they used the terms 'characteristic vector' and 'characteristic value'. These were translated into German as 'Eigenvektor' and 'Eigenwerte', and then partially translated back into English—largely by physicists—as 'eigenvector' and 'eigenvalue'. Some English and American mathematicians tried to retain the original English terms, but they were overwhelmed by extensive use of the physicists' language in applications. Nowadays everyone uses the German terms. The one exception is that we still call

$$\det(A - \lambda I) = 0$$

the characteristic equation and not some strange German-English name.

Solving Polynomial Equations. To find the eigenvalues of an $n \times n$ matrix, you have to solve a polynomial equation. You all know how to solve quadratic equations, but you may be stumped by cubic or higher equations, particularly if there are no obvious ways to factor. You should review what you learned in high school about this subject, but here are a few guidelines to help you.

First, it is not generally possible to find a simple solution in closed form for an algebraic equation. For most equations you might encounter in practice, you would have to use some method to approximate a solution. (Many such methods exist. One you may have learned in your calculus course is *Newton's Method*.) Unfortunately, an approximate solution of the characteristic equation isn't much good for finding the corresponding eigenvectors. After all, the system

$$(A - \lambda I)\mathbf{v} = 0$$

must have rank smaller than n for there to be non-zero solutions $\mathbf{v}$. If you replace the exact value of λ by an approximation, the chances are that the new system will have rank n. Hence, the textbook method we have described for finding eigenvectors won't work. There are in fact many alternative methods for finding eigenvalues and eigenvectors approximately when exact solutions are not available. Whole books are devoted to such methods. (See *Johnson, Riess, and Arnold* or *Noble and Daniel* for some discussion of these matters.)

Fortunately, textbook exercises and examination questions almost always involve characteristic equations for which exact solutions exist, but it is not always obvious what they are. Here is one fact (a consequence of an important result called *Gauss's Lemma*) which helps us find such exact solutions when they exist. Consider an equation of the form

$$\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n = 0$$

where all the coefficients are *integers*. (The characteristic equation of a matrix always has leading coefficient 1 or -1. In the latter case, just imagine you have multiplied through by -1 to apply the method.) Gauss's Lemma tells us that if this equation has any roots which are *rational numbers*, i.e., quotients of integers, then any such root is actually an integer, and, moreover, it must divide the constant term a_n. Hence, the first step in solving such an equation should be checking all possible factors (positive and negative) of the constant term. Once, you know a root r_1, you can divide through by $\lambda - r_1$ to reduce to a lower degree equation. If you know the method of synthetic division, you will find checking the possible roots and the polynomial long division much simpler.

Example 4.4. Solve

$$\lambda^3 - 3\lambda + 2 = 0.$$

If there are any rational roots, they must be factors of the constant term 2. Hence, we must try $1, -1, 2, -2$. Substituting $\lambda = 1$ in the equation yields 0, so it is a root. Dividing $\lambda^3 - 3\lambda + 2$ by $\lambda - 1$ yields

$$\lambda^3 - 3\lambda + 2 = (\lambda - 1)(\lambda^2 + \lambda - 2)$$

and this may be factored further to obtain

$$\lambda^3 - 3\lambda^2 + 2 = (\lambda - 1)(\lambda - 1)(\lambda + 2) = (\lambda - 1)^2(\lambda + 2).$$

Hence, the roots are $\lambda = 1$ which is a double root and $\lambda = -2$.

Complex Roots. A polynomial equation may end up having *complex* roots. This can certainly occur for a characteristic equation.

Example 4.5. Let

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

Its characteristic equation is

$$\det(A - \lambda I) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1 = 0.$$

As you learned in high school algebra, the roots of this equation are $\pm i$ where i is the imaginary square root of -1.

In such a case, we won't have much luck in finding eigenvectors in $\mathbf{R}^n$ for such 'eigenvalues', since solving the appropriate linear equations will yield solutions with non-real, complex entries. It is possible to develop a complete theory based on complex scalars and complex entries, and such a theory is very useful in certain areas like electrical engineering. For the moment, however, we shall restrict our attention to the theory in which everything is assumed to be real. In that context, we just ignore non-real, complex roots of the characteristic equation

Appendix. Proof of the linear independence of sets of eigenvectors for distinct eigenvalues. Assume $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is not a linearly independent set, and try to derive a contradiction. In this case, one of the vectors in the set can be expressed as a linear combination of the others. If we number the elements appropriately, we may assume that

$$\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_r, \tag{30}$$

where $r \le k$. (Before renumbering, leave out any vector $\mathbf{v}_i$ on the right if it appears with coefficient $c_i = 0$.) Note that we may also assume that no vector which appears on the right is a linear combination of the others because otherwise we could express it so and after combining terms delete it from the sum. Thus we may assume the vectors which appear on the right form a linearly independent set. Multiply (30) on the left by A. We get

$$A\mathbf{v}_1 = c_2 A\mathbf{v}_2 + \cdots + c_k A\mathbf{v}_k \tag{31}$$

$$\lambda_1\mathbf{v}_1 = c_2\lambda_2\mathbf{v}_2 + \ldots c_k\lambda_k\mathbf{v}_k \tag{32}$$

where in (31) we used the fact that each $\mathbf{v}_i$ is an eigenvector with eigenvalue λ_i. Now multiply (30) by λ_1 and subtract from (31). We get

$$0 = c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + \cdots + c_k(\lambda_k - \lambda_1)\mathbf{v}_k. \tag{33}$$

Not all the coefficients on the right in this equation are zero. For at least one of the $c_i \ne 0$ (since $\mathbf{v}_1 \ne 0$), and none of the quantities $\lambda_2 - \lambda_1, \ldots \lambda_k - \lambda_1$ is zero. It follows that (31) may be used to express one of the vectors $\mathbf{v}_2, \ldots, \mathbf{v}_k$ as a linear combination of the others. However, this contradicts the assertion that the set of vectors appearing on the right is linearly independent. Hence, our initial assumption that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is dependent must be false, and the theorem is proved.

You should try this argument out on a set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of three eigenvectors to see if you understand it.

Exercises for Section 4.

4.1. Find the eigenvalues and eigenvectors for each of the following matrices. Use the method given in the text for solving the characteristic equation if it has degree greater than two.

(a) $\begin{bmatrix} 5 & -3 \\ 2 & 0 \end{bmatrix}$.

(b) $\begin{bmatrix} 3 & -2 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix}$.

(c) $\begin{bmatrix} 2 & -1 & -1 \\ 0 & 0 & -2 \\ 0 & 1 & 3 \end{bmatrix}$.

(d) $\begin{bmatrix} 4 & -1 & -1 \\ 0 & 2 & -1 \\ 1 & 0 & 3 \end{bmatrix}$.

4.2. You are a mechanical engineer checking for metal fatigue in a vibrating system. Mathematical analysis reduces the problem to finding eigenvectors for the matrix $A = \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$. A member of your design team tells you that $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is an eigenvector for A. What is the quickest way for you to check if this is correct?

4.3. As in the previous problem, some other member of your design team tells you that $\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is a basis for the eigenspace of the same matrix corresponding to one of its eigenvalues. What do you say in return?

4.4. Under what circumstances can zero be an eigenvalue of the square matrix A? Could A be non-singular in this case? Hint: The characteristic equation is $\det(A - \lambda I) = 0$.

4.5. Let A be a square matrix, and suppose λ is an eigenvalue for A with eigenvector $\mathbf{v}$. Show that λ^2 is an eigenvalue for A^2 with eigenvector $\mathbf{v}$. What about λ^n and A^n for n a positive integer?

4.6. Suppose A is non-singular. Show that λ is an eigenvalue of A if and only if λ^{-1} is an eigenvalue of A^{-1}. Hint. Use the same eigenvector.

4.7. (a) Show that $\det(A - \lambda I)$ is a quadratic polynomial in λ if A is a 2×2 matrix.

(b) Show that $\det(A - \lambda I)$ is a cubic polynomial in λ if A is a 3×3 matrix.

(c) What would you guess is the coefficient of λ^n in $\det(A - \lambda I)$ for A an $n \times n$ matrix?

4.8. (Optional) Let A be an $n \times n$ matrix with entries not involving λ. Prove in general that $\det(A - \lambda I)$ is a polynomial in λ of degree n. Hint: Assume $B(\lambda)$ is an $n \times n$ matrix such that each column has at most one term involving λ and that term is of the form $a + b\lambda$. Show by using the recursive definition of the determinant that $\det B(\lambda)$ is a polynomial in λ of degree at most n. Now use this fact and the recursive definition of the determinant to show that $\det(A - \lambda I)$ is a polynomial of degree exactly n.

4.9. (Project) The purpose of this project is to illustrate one method for approximating an eigenvector and the corresponding eigenvalue in cases where exact calculation is not feasible. We use an example in which one can find exact answers by the usual method, at least if one uses radicals, so we can compare answers to gauge how effective the method is.

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. Define an infinite sequence of vectors $\mathbf{v}_n$ in $\mathbf{R}^2$ as follows. Let $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Having defined $\mathbf{v}_n$, define $\mathbf{v}_{n+1} = A\mathbf{v}_n$. Thus, $\mathbf{v}_1 = A\mathbf{v}_0, \mathbf{v}_2 = A\mathbf{v}_1 = A^2\mathbf{v}_0, \mathbf{v}_3 = A\mathbf{v}_2 = A^3\mathbf{v}_0$, etc. Then it turns out in this case that as $n \to \infty$, the *directions* of the vectors $\mathbf{v}_n$ approach the *direction* of an eigenvector for A. Unfortunately, there is one difficulty: the magnitudes $|\mathbf{v}_n|$ approach infinity. To get get around this problem, proceed as follows. Let $\mathbf{v}_n = \begin{bmatrix} a_n \\ b_n \end{bmatrix}$ and put $\mathbf{u}_n = \frac{1}{b_n}\mathbf{v}_n = \begin{bmatrix} a_n/b_n \\ 1 \end{bmatrix}$. Then the second component is always one, and the first component $r_n = a_n/b_n$ approaches a limit r and $\mathbf{u} = \begin{bmatrix} r \\ 1 \end{bmatrix}$ is an eigenvector for A.

(a) For the above matrix A, calculate the sequence of vectors $\mathbf{v}_n$ and and numbers r_n $n = 1, 2, 3, \ldots$. Do the calculations for enough n to see a pattern emerging and so that you can estimate r accurately to 3 decimal places.

(b) Once you know an eigenvector $\mathbf{u}$, you can find the corresponding eigenvalue λ by calculating $A\mathbf{u}$. Use your estimate in part (a) to estimate the corresponding λ.

(c) Compare this to the roots of the characteristic equation $(1-\lambda)(2-\lambda)-1 = 0$. Note that the method employed here only gives you *one* of the two eigenvalues. In fact, this method, when it works, usually gives the largest eigenvalue.

5. Diagonalization

In many cases, the process outlined in the previous section results in a basis for $\mathbf{R}^n$ which consists of eigenvectors for the matrix A. Indeed, the set of eigenvectors so obtained is always linearly independent, so if it is large enough (i.e., has n elements), it will be a basis. When that is the case, the use of that basis to establish a coordinate system for $\mathbf{R}^n$ can simplify calculations involving A.

Example 5.1. Let

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

We found in the previous section that

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

are eigenvectors respectively with eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$. The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent, and since it has two elements, it must be a basis for $\mathbf{R}^2$. Suppose $\mathbf{v}$ is any vector in $\mathbf{R}^2$. We may express it respect to this new basis

$$\mathbf{v} = \mathbf{v}_1 y_1 + \mathbf{v}_2 y_2 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

where (y_1, y_2) are the coordinates of $\mathbf{v}$ with respect to this new basis. It follows that

$$A\mathbf{v} = A(\mathbf{v}_1 y_1 + \mathbf{v}_2 y_2) = (A\mathbf{v}_1)y_1 + (A\mathbf{v}_2)y_2.$$

However, since they are eigenvectors, each is just multiplied by the corresponding eigenvalue, or in symbols

$$A\mathbf{v}_1 = \mathbf{v}_1(3) \quad \text{and} \quad A\mathbf{v}_2 = \mathbf{v}_2(1).$$

So

$$A(\mathbf{v}_1 y_1 + \mathbf{v}_2 y_2) = \mathbf{v}_1(3y_1) + \mathbf{v}_2 y_2 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 3y_1 \\ y_2 \end{bmatrix}. \tag{34}$$

In other words, with respect to the *new coordinates*, the effect of multiplication by A on a vector is to multiply the first new coordinate by the first eigenvalue $\lambda_1 = 3$ and the second new coordinate by the second eigenvalue $\lambda_2 = 1$.

Whenever there is a basis for $\mathbf{R}^n$ consisting of eigenvectors for A, we say that A is *diagonalizable* and that the new basis *diagonalizes* A.

The reason for this terminology may be explained as follows. In the above example, rewrite the left most side of equation (34)

$$A(\mathbf{v}_1 y_1 + \mathbf{v}_2 y_2) = A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{35}$$

and the right most side as

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 3y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \tag{36}$$

These expressions in (35) and (36) are equal for *all* y_1, y_2; in particular, for

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Since these two choices pick out the columns of the matrix, we may drop the common factor $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ on the right, and we get

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Let

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

i.e., P is the 2×2 matrix with the basic eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ as columns. Then, the above equation can be written

$$AP = P \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad P^{-1}AP = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

(The reader should check explicitly in this case that

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.\Bigg)$$

By means of these steps, the matrix A has been expressed in terms of a diagonal matrix with its eigenvalues on the diagonal. This process is called *diagonalization* We shall return in Chapter 3 to a more extensive discussion of diagonalization.

Example 5.2. Not every $n \times n$ matrix A is diagonalizable. That is, it is not always possible to find a basis for $\mathbf{R}^n$ consisting of eigenvectors for A. For example, let

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

The characteristic equation is

$$\det \begin{bmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} = (3-\lambda)^2 = 0.$$

There is only one root $\lambda = 3$ which is a *double* root of the equation. To find the corresponding eigenvectors, we solve the homogeneous system $(A - 3I)\mathbf{v} = \mathbf{0}$. The coefficient matrix

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

is already reduced, and the corresponding system has the general solution

$$v_2 = 0, \qquad v_1 \quad \text{free.}$$

The general solution vector is

$$\mathbf{v} = \begin{bmatrix} v_1 \\ 0 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = v_1 \mathbf{e}_1.$$

Hence, the eigenspace for $\lambda = 3$ is one dimensional with basis $\{\mathbf{e}_1\}$. There are no other eigenvectors except for multiples of $\mathbf{e}_1$. Thus, we can't possibly find a basis for $\mathbf{R}^2$ consisting of eigenvectors for A.

Note how Example 5.2 differs from the examples which preceded it; its characteristic equation has a repeated root. In fact, we have the following general principle.

If the roots of the characteristic equation of a matrix are all distinct, then there is necessarily a basis for $\mathbf{R}^n$ consisting of eigenvectors, and the matrix is diagonalizable.

In general, if the characteristic equation has repeated roots, then the matrix need not be diagonalizable. However, we might be lucky, and such a matrix may still be diagonalizable.

Example 5.3. Consider the matrix

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

First solve the characteristic equation

$$\det \begin{bmatrix} 1-\lambda & 1 & -1 \\ -1 & 3-\lambda & -1 \\ -1 & 1 & 1-\lambda \end{bmatrix} = \tag{37}$$

$$\begin{aligned} &(1-\lambda)((3-\lambda)(1-\lambda)+1) + (1-\lambda+1) - (-1+3-\lambda) \\ &= (1-\lambda)(3-4\lambda+\lambda^2+1) + 2 - \lambda - 2 + \lambda \\ &= (1-\lambda)(\lambda^2-4\lambda+4) \\ &= (1-\lambda)(\lambda-2)^2 = 0. \end{aligned}$$

Note that 2 is a repeated root. We find the eigenvectors for each of these eigenvalues.

For $\lambda = 2$ we need to solve $(A - 2I)\mathbf{v} = 0$.

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

The general solution of the system is $v_1 = v_2 - v_3$ with v_2, v_3 free. The general solution vector for that system is

$$\mathbf{v} = \begin{bmatrix} v_2 - v_3 \\ v_2 \\ v_3 \end{bmatrix} = v_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

The eigenspace is *two* dimensional. Thus, for the eigenvalue $\lambda = 2$ we obtain *two* basic eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},$$

and any eigenvector for $\lambda = 2$ is a non-trivial linear combination of these.

For $\lambda = 1$, we need to solve $(A - I)\mathbf{v} = 0$.

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

The general solution of the system is $v_1 = v_3, v_2 = v_3$ with v_3 free. The general solution vector is

$$\mathbf{v} = \begin{bmatrix} v_3 \\ v_3 \\ v_3 \end{bmatrix} = v_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The eigenspace is one dimensional, and a basic eigenvector for $\lambda = 1$ is

$$\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

It is not hard to check that the set of these basic eigenvectors

$$\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is linearly independent, so it is a basis for $\mathbf{R}^3$.

The matrix is diagonalizable.

In the above example, the reason we ended up with a basis for $\mathbf{R}^3$ consisting of eigenvectors for A was that there were two basic eigenvectors for the double root $\lambda = 2$. In other words, the dimension of the eigenspace was the same as the multiplicity.

THEOREM 5.1. *Let A be an $n \times n$ matrix. The dimension of the eigenspace corresponding to a given eigenvalue is always less than or equal to the multiplicity of that eigenvalue. In particular, if all the roots of the characteristic polynomial are real, then the matrix will be diagonalizable provided, for every eigenvalue, the dimension of the eigenspace is the same as the multiplicity of the eigenvalue. If this fails for at least one eigenvalue, then the matrix won't be diagonalizable.*

Note. If the characteristic polynomial has non-real, complex roots, the matrix also won't be diagonalizable in our sense, since we require all scalars to be real. However, it might still be diagonalizable in the more general theory allowing complex scalars as entries of vectors and matrices.

Exercises for Section 5.

5.1. (a) Find a basis for $\mathbf{R}^2$ consisting of eigenvectors for

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

(b) Let P be the matrix with columns the basis vectors you found in part (a). Check that $P^{-1}AP$ is diagonal with the eigenvalues on the diagonal.

5.2. (a) Find a basis for $\mathbf{R}^3$ consisting of eigenvectors for

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

(b) Let P be the matrix with columns the basis vectors in part (a). Calculate $P^{-1}AP$ and check that it is diagonal with the diagonal entries the eigenvalues you found.

5.3. (a) Find the eigenvalues and eigenvectors for

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

(b) Is A diagonalizable?

5.4. (a) Find a basis for $\mathbf{R}^3$ consisting of eigenvectors for

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

(b) Find a matrix P such that $P^{-1}AP$ is diagonal. Hint: See Problem 5.1.

5.5. Suppose A is a 5×5 matrix with exactly three (real) eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Suppose these have multiplicities m_1, m_2, and m_3 as roots of the characteristic equation. Let d_1, d_2, and d_3 respectively be the dimensions of the eigenspaces for λ_1, λ_2, and λ_3. In each of the following cases, are the given numbers possible, and if so, is A diagonalizable?
(a) $m_1 = 1, d_1 = 1, m_2 = 2, d_2 = 2, m_3 = 2, d_3 = 2$.
(b) $m_1 = 2, d_1 = 1, m_2 = 1, d_2 = 1, m_3 = 2, d_3 = 2$.
(c) $m_1 = 1, d_1 = 2, m_2 = 2, d_2 = 2, m_3 = 2, d_3 = 2$.
(d) $m_1 = 1, d_1 = 1, m_2 = 1, d_2 = 1, m_3 = 1, d_3 = 1$.

5.6. Tell if each of the following matrices is diagonalizable or not.

$$\text{(a)} \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}, \quad \text{(b)} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

6. The Exponential of a Matrix

Recall the series expansion for the exponential function

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This series is specially well behaved. It converges for all possible x.

There are situations in which one would like to make sense of expressions of the form $f(A)$ where $f(x)$ is a well defined function of a scalar variable and A is a square matrix. One way to do this is to try to make a series expansion. We show how to do this for the exponential function.

Define

$$e^A = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

A little explanation is necessary. Each term on the right is an $n \times n$ matrix. If there were only a finite number of such terms, there would be no problem, and the sum would also be an $n \times n$ matrix. In general, however, there are infinitely many terms, and we have to worry about whether it makes sense to add them up.

EXAMPLE 6.1. Let

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

Then

$$\begin{aligned} A^2 &= t^2 \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ A^3 &= t^3 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\ A^4 &= t^4 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ A^5 &= t^5 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\ &\vdots \end{aligned}$$

Hence,

$$\begin{aligned} e^A &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \frac{1}{2} t^2 \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{1}{3!} t^3 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \dots \\ &= \begin{bmatrix} 1 - \frac{t^2}{2} + \frac{t^4}{4!} - \dots & t - \frac{t^3}{3!} + \frac{t^5}{5!} - \dots \\ -t + \frac{t^3}{3!} - \frac{t^5}{5!} + \dots & 1 - \frac{t^2}{2} + \frac{t^4}{4!} - \dots \end{bmatrix} \\ &= \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}. \end{aligned}$$

As in the example, a series of $n \times n$ matrices yields a separate series for each of the n^2 possible entries. We shall say that such a series of matrices converges if the series it yields for each entry converges. With this rule, it is possible to show that the series defining e^A converges for any $n \times n$ matrix A, but the proof is a bit involved. Fortunately, it is often the case that we can avoid worrying about convergence by appropriate trickery. In what follows we shall generally ignore such matters and act as if the series were finite sums.

The exponential function for matrices obeys the usual rules you expect an exponential function to have, but sometimes you have to be careful.

(1) If 0 denotes the $n \times n$ zero matrix, then $e^0 = I$.
(2) The law of exponents holds if the matrices commute, i.e., if B and C are $n \times n$ matrices such that $BC = CB$, then $e^{B+C} = e^B e^C$.
(3) If A is an $n \times n$ constant matrix, then $\frac{d}{dt} e^{At} = A e^{At} = e^{At} A$. (It is worth writing this in both orders because products of matrices don't automatically commute.)

Here are the proofs of these facts.

(1) $e^0 = I + 0 + \frac{1}{2} 0^2 + \dots = I$.

(2) See the Exercises.

(3) Here we act as if the sum were finite (although the argument would work in general if we knew enough about convergence of series of matrices.)

$$\begin{aligned}
\frac{d}{dt}e^{At} &= \frac{d}{dt}\left(I + tA + \frac{1}{2}t^2A^2 + \frac{1}{3!}t^3A^3 + \dots + \frac{1}{j!}t^jA^j + \dots\right) \\
&= 0 + A + \frac{1}{2}(2t)A^2 + \frac{1}{3!}(3t^2)A^3 + \dots + \frac{1}{j!}(jt^{j-1})A^j + \dots \\
&= A + tA^2 + \frac{1}{2}t^2A^3 + \dots + \frac{1}{(j-1)!}t^{j-1}A^j + \dots \\
&= A(I + tA + \frac{1}{2}t^2A^2 + \dots + \frac{1}{(j-1)!}t^{j-1}A^{j-1} + \dots) \\
&= Ae^{At}.
\end{aligned}$$

Note that in the next to last step A could just as well have been factored out on the right, so it doesn't matter which side you put it on.

6.1. Exercises for Section 6.

6.1. (a) Let $A = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$. Show that

$$e^{At} = \begin{bmatrix} e^{\lambda t} & 0 \\ 0 & e^{\mu t} \end{bmatrix}.$$

(b) Let $A = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$. Such a matrix is called a *diagonal matrix*. What can you say about e^{At}?

6.2. (a) Let $N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Calculate e^{Nt}.

(b) Let $N = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Calculate e^{Nt}.

(c) Let N be an $n \times n$ matrix of the form $\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$. What is the smallest integer k satisfying $N^k = 0$? What can you say about e^{Nt}?

6.3. (a) Let $A = \begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$. Calculate e^{At}. Hint: use $A = \lambda I + (A - \lambda I)$. Note that A and $N = A - \lambda I$ commute.

(b) Let $A = \begin{bmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{bmatrix}$. Calculate e^{At}.

(c) Let A be an $n\times n$ matrix of the form $\begin{bmatrix} \lambda & 0 & \dots & 0 & 0 \\ 1 & \lambda & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \lambda \end{bmatrix}$. What can you say about $e^{At} = e^{\lambda t}e^{(A-\lambda I)t}$?

6.4. Let A be an $n\times n$ matrix, and let P be a non-singular $n\times n$ matrix. Show that

$$Pe^{At}P^{-1} = e^{PAP^{-1}t}.$$

6.5. Let B and C be two $n\times n$ matrices such that $BC = CB$. Prove that

$$e^{B+C} = e^B e^C.$$

Hint: You may assume that the binomial theorem applies to commuting matrices, i.e.,

$$(B+C)^n = \sum_{i+j=n} \frac{n!}{i!j!} B^i C^j.$$

6.6. Let

$$B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}.$$

(a) Show that $BC \neq CB$.
(b) Show that

$$e^B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad e^C = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

(c) Show that $e^B e^C \neq e^{B+C}$. Hint: $B + C = J$, where e^{tJ} was calculated in the text.

7. Review

Exercises for Section 7.

7.1. What is $\det \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 1 \\ 3 & 2 & 1 \end{bmatrix}$? Find the answer without using the recursive formula or Gaussian reduction.

7.2. Tell whether each of the following statements is true or false. and, if false, explain why.
(a) If A and B are $n\times n$ matrices then $\det(AB) = \det A \det B$.
(b) If A is an $n\times n$ matrix and c is a scalar, then $\det(cA) = c \det A$.
(c) If A is $m\times n$ and B is $n\times p$, then $(AB)^t = B^t A^t$.
(d) If A is invertible, then so it A^t.

7.3. $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

7.4. Find $\det \begin{bmatrix} 1 & 3 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 2 & 4 & 1 & 1 \\ 1 & 1 & 6 & 1 \end{bmatrix}$.

7.5. Each of the following statements is not generally true. In each case explain briefly why it is false.

(a) An $n \times n$ matrix A is invertible if and only if $\det A = 0$.

(b) If A is an $n \times n$ real matrix, then there is a basis for $\mathbf{R}^n$ consisting of eigenvectors for A.

(c) $\det A^t = \det A$. Hint. Are these defined?

7.6. Let $A = \begin{bmatrix} 2 & 1 & 6 \\ 1 & 3 & 1 \\ 2 & 2 & 5 \end{bmatrix}$. Is $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ an eigenvector for A? Justify your answer.

7.7. (a) The characteristic equation of

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

is $-(\lambda - 3)^2(\lambda - 1) = 0$. Is A diagonalizable? Explain.

(b) Is $B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ diagonalizable? Explain.

7.8. Let A be an $n \times n$ matrix with the property that the sum of all the entries in each row is always the same number a. Without using determinants, show that the common sum a is an eigenvalue. Hint: What is the corresponding eigenvector?

CHAPTER 3

Applications

1. Real Symmetric Matrices

The most common matrices we meet in applications are *symmetric*, that is, they are square matrices which are equal to their transposes. In symbols, $A^t = A$.

EXAMPLES.

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

are symmetric, but

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

are not.

Symmetric matrices are in many ways much simpler to deal with than general matrices.

First, as we noted previously, it is not generally true that the roots of the characteristic equation of a matrix are necessarily real numbers, even if the matrix has only real entries. However,

if A is a symmetric matrix with real entries, then the roots of its characteristic equation are all real.

EXAMPLE 1.1. The characteristic equations of

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

are

$$\lambda^2 - 1 = 0 \qquad \text{and} \qquad \lambda^2 + 1 = 0$$

respectively. In the first case, the roots are real numbers ± 1 while in the second case the roots are imaginary numbers $\pm i$. Notice the dramatic effect of a simple change of sign.

The reason for the reality of the roots (for a real symmetric matrix) is a bit subtle, and we will come back to it later sections.

The second important property of real symmetric matrices is that *they are always diagonalizable*, that is, there is always a basis for $\mathbf{R}^n$ consisting of eigenvectors for the matrix.