3.6 PROPERTIES OF MATRIX MULTIPLICATION

We saw in the previous section that there are some differences between scalar and matrix algebra—most notable is the fact that matrix multiplication is not commutative, and there is no cancellation law. But there are also some important similarities, and the purpose of this section is to look deeper into these issues.

Although we can adjust to not having the commutative property, the situation would be unbearable if the distributive and associative properties were not available. Fortunately, both of these properties hold for matrix multiplication.

Distributive and Associative Laws

For conformable matrices each of the following is true.

- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ (left-hand distributive law).
- $(\mathbf{D}+\mathbf{E})\mathbf{F} = \mathbf{DF}+\mathbf{EF}$ (right-hand distributive law).
- $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ (associative law).

Proof. To prove the left-hand distributive property, demonstrate the corresponding entries in the matrices $\mathbf{A}(\mathbf{B}+\mathbf{C})$ and $\mathbf{AB}+\mathbf{AC}$ are equal. To this end, use the definition of matrix multiplication to write

$$\begin{aligned}
[\mathbf{A}(\mathbf{B}+\mathbf{C})]_{ij} &= \mathbf{A}_{i*}(\mathbf{B}+\mathbf{C})_{*j} = \sum_k [\mathbf{A}]_{ik}[\mathbf{B}+\mathbf{C}]_{kj} = \sum_k [\mathbf{A}]_{ik}\left([\mathbf{B}]_{kj}+[\mathbf{C}]_{kj}\right) \\
&= \sum_k \left([\mathbf{A}]_{ik}[\mathbf{B}]_{kj}+[\mathbf{A}]_{ik}[\mathbf{C}]_{kj}\right) = \sum_k [\mathbf{A}]_{ik}[\mathbf{B}]_{kj} + \sum_k [\mathbf{A}]_{ik}[\mathbf{C}]_{kj} \\
&= \mathbf{A}_{i*}\mathbf{B}_{*j} + \mathbf{A}_{i*}\mathbf{C}_{*j} = [\mathbf{AB}]_{ij} + [\mathbf{AC}]_{ij} \\
&= [\mathbf{AB}+\mathbf{AC}]_{ij}.
\end{aligned}$$

Since this is true for each i and j, it follows that $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$. The proof of the right-hand distributive property is similar and is omitted. To prove the associative law, suppose that $\mathbf{B}$ is $p\times q$ and $\mathbf{C}$ is $q\times n$, and recall from (3.5.7) that the j^{th} column of $\mathbf{BC}$ is a linear combination of the columns in $\mathbf{B}$. That is,

$$[\mathbf{BC}]_{*j} = \mathbf{B}_{*1}c_{1j} + \mathbf{B}_{*2}c_{2j} + \cdots + \mathbf{B}_{*q}c_{qj} = \sum_{k=1}^{q} \mathbf{B}_{*k}c_{kj}.$$

Use this along with the left-hand distributive property to write

$$[\mathbf{A}(\mathbf{BC})]_{ij} = \mathbf{A}_{i*}[\mathbf{BC}]_{*j} = \mathbf{A}_{i*}\sum_{k=1}^{q}\mathbf{B}_{*k}c_{kj} = \sum_{k=1}^{q}\mathbf{A}_{i*}\mathbf{B}_{*k}c_{kj}$$
$$= \sum_{k=1}^{q}[\mathbf{AB}]_{ik}c_{kj} = [\mathbf{AB}]_{i*}\mathbf{C}_{*j} = [(\mathbf{AB})\mathbf{C}]_{ij}. \quad \blacksquare$$

Example 3.6.1

Linearity of Matrix Multiplication. Let $\mathbf{A}$ be an $m \times n$ matrix, and f be the function defined by matrix multiplication

$$f(\mathbf{X}_{n\times p}) = \mathbf{AX}.$$

The left-hand distributive property guarantees that f is a linear function because for all scalars α and for all $n \times p$ matrices $\mathbf{X}$ and $\mathbf{Y}$,

$$f(\alpha\mathbf{X}+\mathbf{Y}) = \mathbf{A}(\alpha\mathbf{X}+\mathbf{Y}) = \mathbf{A}(\alpha\mathbf{X}) + \mathbf{AY} = \alpha\mathbf{AX} + \mathbf{AY}$$
$$= \alpha f(\mathbf{X}) + f(\mathbf{Y}).$$

Of course, the linearity of matrix multiplication is no surprise because it was the consideration of linear functions that motivated the definition of the matrix product at the outset.

For scalars, the number 1 is the identity element for multiplication because it has the property that it reproduces whatever it is multiplied by. For matrices, there is an identity element with similar properties.

Identity Matrix

The $n \times n$ matrix with 1's on the main diagonal and 0's elsewhere

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is called the ***identity matrix*** of order n. For every $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{AI}_n = \mathbf{A} \quad \text{and} \quad \mathbf{I}_m\mathbf{A} = \mathbf{A}.$$

The subscript on $\mathbf{I}_n$ is neglected whenever the size is obvious from the context.

Proof. Notice that $\mathbf{I}_{*j}$ has a 1 in the j^{th} position and 0's elsewhere. Recall from Exercise 3.5.4 that such columns were called ***unit columns,*** and they have the property that for any conformable matrix $\mathbf{A}$,

$$\mathbf{AI}_{*j} = \mathbf{A}_{*j}.$$

Using this together with the fact that $[\mathbf{AI}]_{*j} = \mathbf{AI}_{*j}$ produces

$$\mathbf{AI} = \begin{pmatrix} \mathbf{AI}_{*1} & \mathbf{AI}_{*2} & \cdots & \mathbf{AI}_{*n} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{*1} & \mathbf{A}_{*2} & \cdots & \mathbf{A}_{*n} \end{pmatrix} = \mathbf{A}.$$

A similar argument holds when $\mathbf{I}$ appears on the left-hand side of $\mathbf{A}$. ■

Analogous to scalar algebra, we define the 0^{th} power of a square matrix to be the identity matrix of corresponding size. That is, if $\mathbf{A}$ is $n \times n$, then

$$\mathbf{A}^0 = \mathbf{I}_n.$$

Positive powers of $\mathbf{A}$ are also defined in the natural way. That is,

$$\mathbf{A}^n = \underbrace{\mathbf{AA}\cdots\mathbf{A}}_{n \text{ times}}.$$

The associative law guarantees that it makes no difference how matrices are grouped for powering. For example, $\mathbf{AA}^2$ is the same as $\mathbf{A}^2\mathbf{A}$, so that

$$\mathbf{A}^3 = \mathbf{AAA} = \mathbf{AA}^2 = \mathbf{A}^2\mathbf{A}.$$

Also, the usual laws of exponents hold. For nonnegative integers r and s,

$$\mathbf{A}^r\mathbf{A}^s = \mathbf{A}^{r+s} \quad \text{and} \quad (\mathbf{A}^r)^s = \mathbf{A}^{rs}.$$

We are not yet in a position to define negative or fractional powers, and due to the lack of conformability, powers of nonsquare matrices are never defined.

Example 3.6.2

A Pitfall. For two $n \times n$ matrices, what is $(\mathbf{A}+\mathbf{B})^2$? **Be careful!** Because matrix multiplication is not commutative, the familiar formula from scalar algebra is not valid for matrices. The distributive properties must be used to write

$$\begin{aligned}(\mathbf{A}+\mathbf{B})^2 &= \underbrace{(\mathbf{A}+\mathbf{B})}(\mathbf{A}+\mathbf{B}) = \underbrace{(\mathbf{A}+\mathbf{B})}\mathbf{A} + \underbrace{(\mathbf{A}+\mathbf{B})}\mathbf{B} \\ &= \mathbf{A}^2 + \mathbf{BA} + \mathbf{AB} + \mathbf{B}^2,\end{aligned}$$

and this is as far as you can go. The familiar form $\mathbf{A}^2+2\mathbf{AB}+\mathbf{B}^2$ is obtained only in those rare cases where $\mathbf{AB} = \mathbf{BA}$. To evaluate $(\mathbf{A}+\mathbf{B})^k$, the distributive rules must be applied repeatedly, and the results are a bit more complicated—try it for $k = 3$.

Example 3.6.3

Suppose that the population migration between two geographical regions—say, the North and the South—is as follows. Each year, 50% of the population in the North migrates to the South, while only 25% of the population in the South moves to the North. This situation is depicted by drawing a transition diagram such as that shown in Figure 3.6.1.

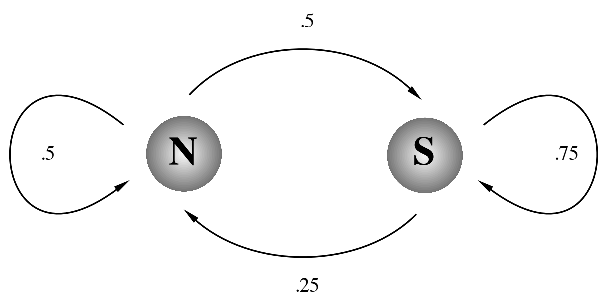

FIGURE 3.6.1

Problem: If this migration pattern continues, will the population in the North continually shrink until the entire population is eventually in the South, or will the population distribution somehow stabilize before the North is completely deserted?

Solution: Let n_k and s_k denote the respective proportions of the total population living in the North and South at the end of year k and assume $n_k + s_k = 1$. The migration pattern dictates that the fractions of the population in each region at the end of year $k+1$ are

$$\begin{aligned} n_{k+1} &= n_k(.5) + s_k(.25), \\ s_{k+1} &= n_k(.5) + s_k(.75). \end{aligned} \tag{3.6.1}$$

If $\mathbf{p}_k^T = (n_k, s_k)$ and $\mathbf{p}_{k+1}^T = (n_{k+1}, s_{k+1})$ denote the respective population distributions at the end of years k and $k+1$, and if

$$\mathbf{T} = \begin{array}{c} \\ \mathrm{N} \\ \mathrm{S} \end{array} \!\!\begin{array}{c} \begin{array}{cc} \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} .5 & .5 \\ .25 & .75 \end{pmatrix} \end{array}$$

is the associated ***transition matrix,*** then (3.6.1) assumes the matrix form $\mathbf{p}_{k+1}^T = \mathbf{p}_k^T\mathbf{T}$. Inducting on $\mathbf{p}_1^T = \mathbf{p}_0^T\mathbf{T}$, $\mathbf{p}_2^T = \mathbf{p}_1^T\mathbf{T} = \mathbf{p}_0^T\mathbf{T}^2$, $\mathbf{p}_3^T = \mathbf{p}_2^T\mathbf{T} = \mathbf{p}_0^T\mathbf{T}^3$, etc., leads to

$$\mathbf{p}_k^T = \mathbf{p}_0^T\mathbf{T}^k. \tag{3.6.2}$$

Determining the long-run behavior involves evaluating $\lim_{k\to\infty} \mathbf{p}_k^T$, and it's clear from (3.6.2) that this boils down to analyzing $\lim_{k\to\infty} \mathbf{T}^k$. Later, in Example

7.3.5, a more sophisticated approach is discussed, but for now we will use the "brute force" method of successively powering $\mathbf{P}$ until a pattern emerges. The first several powers of $\mathbf{P}$ are shown below with three significant digits displayed.

$$\mathbf{P}^2 = \begin{pmatrix} .375 & .625 \\ .312 & .687 \end{pmatrix} \quad \mathbf{P}^3 = \begin{pmatrix} .344 & .656 \\ .328 & .672 \end{pmatrix} \quad \mathbf{P}^4 = \begin{pmatrix} .328 & .672 \\ .332 & .668 \end{pmatrix}$$

$$\mathbf{P}^5 = \begin{pmatrix} .334 & .666 \\ .333 & .667 \end{pmatrix} \quad \mathbf{P}^6 = \begin{pmatrix} .333 & .667 \\ .333 & .667 \end{pmatrix} \quad \mathbf{P}^7 = \begin{pmatrix} .333 & .667 \\ .333 & .667 \end{pmatrix}$$

This sequence appears to be converging to a limiting matrix of the form

$$\mathbf{P}^\infty = \lim_{k\to\infty} \mathbf{P}^k = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix},$$

so the limiting population distribution is

$$\begin{aligned} \mathbf{p}_\infty^T &= \lim_{k\to\infty} \mathbf{p}_k^T = \lim_{k\to\infty} \mathbf{p}_0^T \mathbf{T}^k = \mathbf{p}_0^T \lim_{k\to\infty} \mathbf{T}^k = \begin{pmatrix} n_0 & s_0 \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{n_0+s_0}{3} & \dfrac{2(n_0+s_0)}{3} \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \end{pmatrix}. \end{aligned}$$

Therefore, if the migration pattern continues to hold, then the population distribution will eventually stabilize with 1/3 of the population being in the North and 2/3 of the population in the South. And this is independent of the initial distribution! The powers of $\mathbf{P}$ indicate that the population distribution will be practically stable in no more than 6 years—individuals may continue to move, but the proportions in each region are essentially constant by the sixth year.

The operation of transposition has an interesting effect upon a matrix product—a reversal of order occurs.

Reverse Order Law for Transposition

For conformable matrices $\mathbf{A}$ and $\mathbf{B}$,

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

The case of conjugate transposition is similar. That is,

$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*.$$

Proof. By definition,

$$(\mathbf{AB})^T_{ij} = [\mathbf{AB}]_{ji} = \mathbf{A}_{j*}\mathbf{B}_{*i}.$$

Consider the (i,j)-entry of the matrix $\mathbf{B}^T\mathbf{A}^T$ and write

$$\begin{aligned}\left[\mathbf{B}^T\mathbf{A}^T\right]_{ij} &= \left(\mathbf{B}^T\right)_{i*}\left(\mathbf{A}^T\right)_{*j} = \sum_k \left[\mathbf{B}^T\right]_{ik}\left[\mathbf{A}^T\right]_{kj} \\ &= \sum_k [\mathbf{B}]_{ki}[\mathbf{A}]_{jk} = \sum_k [\mathbf{A}]_{jk}[\mathbf{B}]_{ki} \\ &= \mathbf{A}_{j*}\mathbf{B}_{*i}.\end{aligned}$$

Therefore, $(\mathbf{AB})^T_{ij} = \left[\mathbf{B}^T\mathbf{A}^T\right]_{ij}$ for all i and j, and thus $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. The proof for the conjugate transpose case is similar. ∎

Example 3.6.4

For every matrix $\mathbf{A}_{m\times n}$, the products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are symmetric matrices because

$$\left(\mathbf{A}^T\mathbf{A}\right)^T = \mathbf{A}^T\mathbf{A}^{T^T} = \mathbf{A}^T\mathbf{A} \quad \text{and} \quad \left(\mathbf{A}\mathbf{A}^T\right)^T = \mathbf{A}^{T^T}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.$$

Example 3.6.5

Trace of a Product. Recall from Example 3.3.1 that the trace of a square matrix is the sum of its main diagonal entries. Although matrix multiplication is not commutative, the trace function is one of the few cases where the order of the matrices can be changed without affecting the results.

Problem: For matrices $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{n\times m}$, prove that

$$\mathit{trace}\,(\mathbf{AB}) = \mathit{trace}\,(\mathbf{BA}).$$

Solution:

$$\begin{aligned}\mathit{trace}\,(\mathbf{AB}) &= \sum_i [\mathbf{AB}]_{ii} = \sum_i \mathbf{A}_{i*}\mathbf{B}_{*i} = \sum_i\sum_k a_{ik}b_{ki} = \sum_i\sum_k b_{ki}a_{ik} \\ &= \sum_k\sum_i b_{ki}a_{ik} = \sum_k \mathbf{B}_{k*}\mathbf{A}_{*k} = \sum_k [\mathbf{BA}]_{kk} = \mathit{trace}\,(\mathbf{BA}).\end{aligned}$$

Note: This is true in spite of the fact that $\mathbf{AB}$ is $m\times m$ while $\mathbf{BA}$ is $n\times n$. Furthermore, this result can be extended to say that any product of conformable matrices can be permuted *cyclically* without altering the trace of the product. For example,

$$\mathit{trace}\,(\mathbf{ABC}) = \mathit{trace}\,(\mathbf{BCA}) = \mathit{trace}\,(\mathbf{CAB}).$$

However, a noncyclical permutation may not preserve the trace. For example,

$$\mathit{trace}\,(\mathbf{ABC}) \neq \mathit{trace}\,(\mathbf{BAC}).$$

Executing multiplication between two matrices by partitioning one or both factors into ***submatrices***—a matrix contained within another matrix—can be a useful technique.

Block Matrix Multiplication

Suppose that $\mathbf{A}$ and $\mathbf{B}$ are partitioned into submatrices—often referred to as ***blocks***—as indicated below.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1r} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{s1} & \mathbf{A}_{s2} & \cdots & \mathbf{A}_{sr} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1t} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{r1} & \mathbf{B}_{r2} & \cdots & \mathbf{B}_{rt} \end{pmatrix}.$$

If the pairs $(\mathbf{A}_{ik}, \mathbf{B}_{kj})$ are conformable, then $\mathbf{A}$ and $\mathbf{B}$ are said to be ***conformably partitioned***. For such matrices, the product $\mathbf{AB}$ is formed by combining the blocks exactly the same way as the scalars are combined in ordinary matrix multiplication. That is, the (i, j)-block in $\mathbf{AB}$ is

$$\mathbf{A}_{i1}\mathbf{B}_{1j} + \mathbf{A}_{i2}\mathbf{B}_{2j} + \cdots + \mathbf{A}_{ir}\mathbf{B}_{rj}.$$

Although a completely general proof is possible, looking at some examples better serves the purpose of understanding this technique.

Example 3.6.6

Block multiplication is particularly useful when there are patterns in the matrices to be multiplied. Consider the partitioned matrices

$$\mathbf{A} = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) = \begin{pmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}, \quad \mathbf{B} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 2 & 1 & 2 \\ 3 & 4 & 3 & 4 \end{array}\right) = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{C} \end{pmatrix},$$

where

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Using block multiplication, the product $\mathbf{AB}$ is easily computed to be

$$\mathbf{AB} = \begin{pmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{C} \end{pmatrix} = \begin{pmatrix} 2\mathbf{C} & \mathbf{C} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} = \left(\begin{array}{cc|cc} 2 & 4 & 1 & 2 \\ 6 & 8 & 3 & 4 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right).$$

Example 3.6.7

Reducibility. Suppose that $\mathbf{T}_{n\times n}\mathbf{x} = \mathbf{b}$ represents a system of linear equations in which the coefficient matrix is ***block triangular***. That is, $\mathbf{T}$ can be partitioned as

$$\mathbf{T} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \quad \text{where} \quad \mathbf{A} \text{ is } r\times r \text{ and } \mathbf{C} \text{ is } n-r\times n-r. \tag{3.6.3}$$

If $\mathbf{x}$ and $\mathbf{b}$ are similarly partitioned as $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$, then block multiplication shows that $\mathbf{Tx} = \mathbf{b}$ reduces to two smaller systems

$$\begin{aligned} \mathbf{Ax}_1 + \mathbf{Bx}_2 &= \mathbf{b}_1, \\ \mathbf{Cx}_2 &= \mathbf{b}_2, \end{aligned}$$

so if all systems are consistent, a block version of back substitution is possible—i.e., solve $\mathbf{Cx}_2 = \mathbf{b}_2$ for $\mathbf{x}_2$, and substituted this back into $\mathbf{Ax}_1 = \mathbf{b}_1 - \mathbf{Bx}_2$, which is then solved for $\mathbf{x}_1$. For obvious reasons, block-triangular systems of this type are sometimes referred to as ***reducible systems,*** and $\mathbf{T}$ is said to be a ***reducible matrix.*** Recall that applying Gaussian elimination with back substitution to an $n\times n$ system requires about $n^3/3$ multiplications/divisions and about $n^3/3$ additions/subtractions. This means that it's more efficient to solve two smaller subsystems than to solve one large main system. For example, suppose the matrix $\mathbf{T}$ in (3.6.3) is 100×100 while $\mathbf{A}$ and $\mathbf{C}$ are each 50×50. If $\mathbf{Tx} = \mathbf{b}$ is solved without taking advantage of its reducibility, then about $10^6/3$ multiplications/divisions are needed. But by taking advantage of the reducibility, only about $(250\times 10^3)/3$ multiplications/divisions are needed to solve both 50×50 subsystems. Another advantage of reducibility is realized when a computer's main memory capacity is not large enough to store the entire coefficient matrix but is large enough to hold the submatrices.

Exercises for section 3.6

3.6.1. For the partitioned matrices

$$\mathbf{A} = \left(\begin{array}{c|cc|ccc} 1 & 0 & 0 & 3 & 3 & 3 \\ 1 & 0 & 0 & 3 & 3 & 3 \\ \hline 1 & 2 & 2 & 0 & 0 & 0 \end{array}\right) \quad \text{and} \quad \mathbf{B} = \left(\begin{array}{cc} -1 & -1 \\ \hline 0 & 0 \\ 0 & 0 \\ \hline -1 & -2 \\ -1 & -2 \\ -1 & -2 \end{array}\right),$$

use block multiplication with the indicated partitions to form the product $\mathbf{AB}$.

3.6.2. For all matrices $\mathbf{A}_{n\times k}$ and $\mathbf{B}_{k\times n}$, show that the block matrix

$$\mathbf{L} = \begin{pmatrix} \mathbf{I} - \mathbf{BA} & \mathbf{B} \\ 2\mathbf{A} - \mathbf{ABA} & \mathbf{AB} - \mathbf{I} \end{pmatrix}$$

has the property $\mathbf{L}^2 = \mathbf{I}$. Matrices with this property are said to be ***involutory,*** and they occur in the science of cryptography.

3.6.3. For the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix},$$

determine $\mathbf{A}^{300}$. **Hint:** A square matrix $\mathbf{C}$ is said to be ***idempotent*** when it has the property that $\mathbf{C}^2 = \mathbf{C}$. Make use of idempotent submatrices in $\mathbf{A}$.

3.6.4. For every matrix $\mathbf{A}_{m\times n}$, demonstrate that the products $\mathbf{A}^*\mathbf{A}$ and $\mathbf{AA}^*$ are hermitian matrices.

3.6.5. If $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices that commute, prove that the product $\mathbf{AB}$ is also symmetric. If $\mathbf{AB} \neq \mathbf{BA}$, is $\mathbf{AB}$ necessarily symmetric?

3.6.6. Prove that the right-hand distributive property is true.

3.6.7. For each matrix $\mathbf{A}_{n\times n}$, explain why it is impossible to find a solution for $\mathbf{X}_{n\times n}$ in the matrix equation

$$\mathbf{AX} - \mathbf{XA} = \mathbf{I}.$$

Hint: Consider the trace function.

3.6.8. Let $\mathbf{y}^T_{1\times m}$ be a row of unknowns, and let $\mathbf{A}_{m\times n}$ and $\mathbf{b}^T_{1\times n}$ be known matrices.

(a) Explain why the matrix equation $\mathbf{y}^T\mathbf{A} = \mathbf{b}^T$ represents a system of n linear equations in m unknowns.

(b) How are the solutions for $\mathbf{y}^T$ in $\mathbf{y}^T\mathbf{A} = \mathbf{b}^T$ related to the solutions for $\mathbf{x}$ in $\mathbf{A}^T\mathbf{x} = \mathbf{b}$?

3.6.9. A particular electronic device consists of a collection of switching circuits that can be either in an ON state or an OFF state. These electronic switches are allowed to change state at regular time intervals called *clock cycles.* Suppose that at the end of each clock cycle, 30% of the switches currently in the OFF state change to ON, while 90% of those in the ON state revert to the OFF state.

(a) Show that the device approaches an equilibrium in the sense that the proportion of switches in each state eventually becomes constant, and determine these equilibrium proportions.

(b) Independent of the initial proportions, about how many clock cycles does it take for the device to become essentially stable?

3.6.10. Write the following system in the form $\mathbf{T}_{n\times n}\mathbf{x} = \mathbf{b}$, where $\mathbf{T}$ is block triangular, and then obtain the solution by solving two small systems as described in Example 3.6.7.

$$\begin{aligned} x_1 + x_2 + 3x_3 + 4x_4 &= -1, \\ 2x_3 + 3x_4 &= 3, \\ x_1 + 2x_2 + 5x_3 + 6x_4 &= -2, \\ x_3 + 2x_4 &= 4. \end{aligned}$$

3.6.11. Prove that each of the following statements is true for conformable matrices.

(a) $trace\,(\mathbf{ABC}) = trace\,(\mathbf{BCA}) = trace\,(\mathbf{CAB})$.

(b) $trace\,(\mathbf{ABC})$ can be different from $trace\,(\mathbf{BAC})$.

(c) $trace\left(\mathbf{A}^T\mathbf{B}\right) = trace\left(\mathbf{AB}^T\right)$.

3.6.12. Suppose that $\mathbf{A}_{m\times n}$ and $\mathbf{x}_{n\times 1}$ have real entries.

(a) Prove that $\mathbf{x}^T\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

(b) Prove that $trace\left(\mathbf{A}^T\mathbf{A}\right) = 0$ if and only if $\mathbf{A} = \mathbf{0}$.

Solutions for exercises in section 3. 6

3.6.1.

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 + \mathbf{A}_{13}\mathbf{B}_3 \\ \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 + \mathbf{A}_{23}\mathbf{B}_3 \end{pmatrix}$$

$$= \left(\begin{array}{cc} -10 & -19 \\ -10 & -19 \\ \hline -1 & -1 \end{array}\right)$$

3.6.2. Use block multiplication to verify $\mathbf{L}^2 = \mathbf{I}$—be careful not to commute any of the terms when forming the various products.

3.6.3. Partition the matrix as $\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}$, where $\mathbf{C} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ and observe that $\mathbf{C}^2 = \mathbf{C}$. Use this together with block multiplication to conclude that

$$\mathbf{A}^k = \begin{pmatrix} \mathbf{I} & \mathbf{C} + \mathbf{C}^2 + \mathbf{C}^3 + \cdots + \mathbf{C}^k \\ \mathbf{0} & \mathbf{C}^k \end{pmatrix} = \begin{pmatrix} \mathbf{I} & k\mathbf{C} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}.$$

Therefore, $\mathbf{A}^{300} = \begin{pmatrix} 1 & 0 & 0 & 100 & 100 & 100 \\ 0 & 1 & 0 & 100 & 100 & 100 \\ 0 & 0 & 1 & 100 & 100 & 100 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}.$

3.6.4. $(\mathbf{A}^*\mathbf{A})^* = \mathbf{A}^*\mathbf{A}^{**} = \mathbf{A}^*\mathbf{A}$ and $(\mathbf{A}\mathbf{A}^*)^* = \mathbf{A}^{**}\mathbf{A}^* = \mathbf{A}\mathbf{A}^*$.

3.6.5. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T = \mathbf{BA} = \mathbf{AB}$. It is easy to construct a 2×2 example to show that this need not be true when $\mathbf{AB} \neq \mathbf{BA}$.

3.6.6.

$$\begin{aligned} [(\mathbf{D}+\mathbf{E})\mathbf{F}]_{ij} &= (\mathbf{D}+\mathbf{E})_{i*}\mathbf{F}_{*j} = \sum_k [\mathbf{D}+\mathbf{E}]_{ik}[\mathbf{F}]_{kj} = \sum_k \left([\mathbf{D}]_{ik} + [\mathbf{E}]_{ik}\right)[\mathbf{F}]_{kj} \\ &= \sum_k \left([\mathbf{D}]_{ik}[\mathbf{F}]_{kj} + [\mathbf{E}]_{ik}[\mathbf{F}]_{kj}\right) = \sum_k [\mathbf{D}]_{ik}[\mathbf{F}]_{kj} + \sum_k [\mathbf{E}]_{ik}[\mathbf{F}]_{kj} \\ &= \mathbf{D}_{i*}\mathbf{F}_{*j} + \mathbf{E}_{i*}\mathbf{F}_{*j} = [\mathbf{DF}]_{ij} + [\mathbf{EF}]_{ij} \\ &= [\mathbf{DF} + \mathbf{EF}]_{ij}. \end{aligned}$$

3.6.7. If a matrix $\mathbf{X}$ did indeed exist, then

$$\begin{aligned} \mathbf{I} = \mathbf{AX} - \mathbf{XA} \quad &\Longrightarrow \quad trace\,(\mathbf{I}) = trace\,(\mathbf{AX} - \mathbf{XA}) \\ &\Longrightarrow \quad n = trace\,(\mathbf{AX}) - trace\,(\mathbf{XA}) = 0, \end{aligned}$$

which is impossible.

3.6.8. (a) $\mathbf{y}^T\mathbf{A} = \mathbf{b}^T \implies (\mathbf{y}^T\mathbf{A})^T = \mathbf{b}^{T^T} \implies \mathbf{A}^T\mathbf{y} = \mathbf{b}$. This is an $n \times m$ system of equations whose coefficient matrix is $\mathbf{A}^T$. (b) They are the same.

3.6.9. Draw a transition diagram similar to that in Example 3.6.3 with North and South replaced by ON and OFF, respectively. Let x_k be the proportion of switches in the ON state, and let y_k be the proportion of switches in the OFF state after k clock cycles have elapsed. According to the given information,

$$x_k = x_{k-1}(.1) + y_{k-1}(.3)$$
$$y_k = x_{k-1}(.9) + y_{k-1}(.7)$$

so that $\mathbf{p}_k = \mathbf{p}_{k-1}\mathbf{P}$, where

$$\mathbf{p}_k = (\,x_k \quad y_k\,) \quad \text{and} \quad \mathbf{P} = \begin{pmatrix} .1 & .9 \\ .3 & .7 \end{pmatrix}.$$

Just as in Example 3.6.3, $\mathbf{p}_k = \mathbf{p}_0\mathbf{P}^k$. Compute a few powers of $\mathbf{P}$ to find

$$\mathbf{P}^2 = \begin{pmatrix} .280 & .720 \\ .240 & .760 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} .244 & .756 \\ .252 & .748 \end{pmatrix}$$
$$\mathbf{P}^4 = \begin{pmatrix} .251 & .749 \\ .250 & .750 \end{pmatrix}, \quad \mathbf{P}^5 = \begin{pmatrix} .250 & .750 \\ .250 & .750 \end{pmatrix}$$

and deduce that $\mathbf{P}^\infty = \lim_{k\to\infty} \mathbf{P}^k = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix}$. Thus

$$\mathbf{p}_k \to \mathbf{p}_0\mathbf{P}^\infty = \left(\,\tfrac{1}{4}(x_0+y_0) \quad \tfrac{3}{4}(x_0+y_0)\,\right) = \left(\,\tfrac{1}{4} \quad \tfrac{3}{4}\,\right).$$

For practical purposes, the device can be considered to be in equilibrium after about 5 clock cycles—regardless of the initial proportions.

3.6.10. $(\,-4 \quad 1 \quad -6 \quad 5\,)$

3.6.11. (a) $\mathit{trace}\,(\mathbf{ABC}) = \mathit{trace}\,(\mathbf{A}\{\mathbf{BC}\}) = \mathit{trace}\,(\{\mathbf{BC}\}\mathbf{A}) = \mathit{trace}\,(\mathbf{BCA})$. The other equality is similar. (b) Use almost any set of 2×2 matrices to construct an example that shows equality need not hold. (c) Use the fact that $\mathit{trace}\left(\mathbf{C}^T\right) = \mathit{trace}\,(\mathbf{C})$ for all square matrices to conclude that

$$\begin{aligned} \mathit{trace}\left(\mathbf{A}^T\mathbf{B}\right) &= \mathit{trace}\left(\left(\mathbf{A}^T\mathbf{B}\right)^T\right) = \mathit{trace}\left(\mathbf{B}^T\mathbf{A}^{T^T}\right) \\ &= \mathit{trace}\left(\mathbf{B}^T\mathbf{A}\right) = \mathit{trace}\left(\mathbf{A}\mathbf{B}^T\right). \end{aligned}$$

3.6.12. (a) $\mathbf{x}^T\mathbf{x} = 0 \iff \sum_{k=1}^n x_i^2 = 0 \iff x_i = 0$ for each $i \iff \mathbf{x} = \mathbf{0}$.

(b)
$$\begin{aligned} \mathit{trace}\left(\mathbf{A}^T\mathbf{A}\right) = 0 &\iff \sum_i [\mathbf{A}^T\mathbf{A}]_{ii} = 0 \iff \sum_i (\mathbf{A}^T)_{i*}\mathbf{A}_{*i} = 0 \\ &\iff \sum_i \sum_k [\mathbf{A}^T]_{ik}[\mathbf{A}]_{ki} = 0 \iff \sum_i \sum_k [\mathbf{A}]_{ki}[\mathbf{A}]_{ki} = 0 \\ &\iff \sum_i \sum_k [\mathbf{A}]_{ki}^2 = 0 \\ &\iff [\mathbf{A}]_{ki} = 0 \text{ for each } k \text{ and } i \iff \mathbf{A} = \mathbf{0} \end{aligned}$$