6.2 Linear Independence

WHAT RELATION EXTENDS BEING PARALLEL FROM A PAIR TO AN ARBITRARY COLLECTION OF VECTORS? The essential insight is to focus on the linear combinations that are equal to the zero vector.

6.2.0 Definition. Let m and n be positive integers. The vectors $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2, \dots, \vec{\boldsymbol{a}}_n \in \mathbb{K}^m$ are ***linearly dependent*** if there exists scalars $c_1, c_2 \dots, c_n \in \mathbb{K}$, not all zero, such that $c_1\, \vec{\boldsymbol{a}}_1 + c_2\, \vec{\boldsymbol{a}}_2 + \cdots + c_n\, \vec{\boldsymbol{a}}_n = \vec{\boldsymbol{0}}$. Conversely, the vectors $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2, \dots, \vec{\boldsymbol{a}}_n$ are ***linearly independent*** if the vector equation $c_1\, \vec{\boldsymbol{a}}_1 + c_2\, \vec{\boldsymbol{a}}_2 + \cdots + c_n\, \vec{\boldsymbol{a}}_n = \vec{\boldsymbol{0}}$ implies that $c_1 = c_2 = \cdots = c_n = 0$.

To decide whether a collection of vectors is linear dependent or linearly independent, we solve a homogeneous linear system.

6.2.1 Problem. Consider $\vec{\boldsymbol{a}}_1 := \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^{\mathsf{T}}$, $\vec{\boldsymbol{a}}_2 := \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}^{\mathsf{T}}$, and $\vec{\boldsymbol{a}}_3 := \begin{bmatrix} 2 & 1 & 0 \end{bmatrix}^{\mathsf{T}}$. Are the vectors $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2, \vec{\boldsymbol{a}}_3$ linearly independent?

Solution. Set $\mathbf{A} := [\vec{\boldsymbol{a}}_1 \;\; \vec{\boldsymbol{a}}_2 \;\; \vec{\boldsymbol{a}}_3]$ and solve the homogeneous linear system $\mathbf{A}\,\vec{\boldsymbol{x}} = \vec{\boldsymbol{0}}$. The row reduction algorithm [4.2.0] gives

$$\begin{bmatrix} \underline{1} & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \xrightarrow[\sim]{\substack{\vec{r}_2 - 2\vec{r}_1 \mapsto \vec{r}_2 \\ \vec{r}_3 - 3\vec{r}_1 \mapsto \vec{r}_3}} \begin{bmatrix} 1 & 4 & 2 \\ 0 & \underline{-3} & -3 \\ 0 & -6 & -6 \end{bmatrix} \xrightarrow[\sim]{-\frac{1}{3}\vec{r}_2 \mapsto \vec{r}_2} \begin{bmatrix} 1 & 4 & 2 \\ 0 & \underline{1} & 1 \\ 0 & -6 & -6 \end{bmatrix} \xrightarrow[\sim]{\substack{\vec{r}_1 - 4\vec{r}_2 \mapsto \vec{r}_1 \\ \vec{r}_3 + 6\vec{r}_2 \mapsto \vec{r}_3}} \begin{bmatrix} \underline{1} & 0 & -2 \\ 0 & \underline{1} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the solution set is $\operatorname{Span}\big(\begin{bmatrix} 2 & -1 & \underline{1} \end{bmatrix}^{\mathsf{T}}\big)$, the nonzero linear relation $2\,\vec{\boldsymbol{a}}_1 - \vec{\boldsymbol{a}}_2 + \vec{\boldsymbol{a}}_3 = \vec{\boldsymbol{0}}$ certifies that the vectors are linearly dependent. □

6.2.2 Problem. Prove that a single vector $\vec{\boldsymbol{a}} \in \mathbb{K}^m$ is linear independent if and only if it is nonzero.

Solution. The set $\{\vec{\boldsymbol{a}}\}$ is linearly dependent if and only if there exists a nonzero scalar $c \in \mathbb{K}$ such that $c\,\vec{\boldsymbol{a}} = \vec{\boldsymbol{0}}$. It follows that $c\,a_j = 0$, for all $1 \leqslant j \leqslant m$, so we obtain $a_j = c^{-1}(c\,a_j) = c^{-1}(0) = 0$ and $\vec{\boldsymbol{a}} = \vec{\boldsymbol{0}}$. □

6.2.3 Problem. Show that the two vectors $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2 \in \mathbb{K}^m$ are linearly dependent if and only if we have $\vec{\boldsymbol{a}}_1 = \vec{\boldsymbol{0}}$ or $\vec{\boldsymbol{a}}_2$ is a multiple of $\vec{\boldsymbol{a}}_1$.

Solution.

$\Rightarrow$: Suppose that $c_1\,\vec{\boldsymbol{a}}_1 + c_2\,\vec{\boldsymbol{a}}_2 = \vec{\boldsymbol{0}}$ where at least one of the scalars $c_1, c_2 \in \mathbb{K}$ is nonzero. When $c_2 \neq 0$, we have $\vec{\boldsymbol{a}}_2 = \frac{c_1}{c_2}\vec{\boldsymbol{a}}_1$. If $c_2 = 0$, then we must have $c_1 \neq 0$ and $\vec{\boldsymbol{a}}_1 = \vec{\boldsymbol{0}}$.

$\Leftarrow$: When $\vec{\boldsymbol{a}}_1 = \vec{\boldsymbol{0}}$, the vector equation $1\,\vec{\boldsymbol{a}}_1 + 0\,\vec{\boldsymbol{a}}_2 = \vec{\boldsymbol{0}}$ certifies that the vectors are linear dependent. Similarly, when $\vec{\boldsymbol{a}}_2 = c\,\vec{\boldsymbol{a}}_1$, the vector equation $c\,\vec{\boldsymbol{a}}_1 - \vec{\boldsymbol{a}}_2 = \vec{\boldsymbol{0}}$ certifies linear dependence. □

As a counterpart to our universal consistency proposition [6.1.0], we identify homogeneous linear systems with a unique solution.

6.2.4 Proposition (Characterizations of a unique solution). *Fix positive integers m and n. For any $(m \times n)$-matrix $\mathbf{A}$, the following are equivalent.*

a. *The homogeneous linear system $\mathbf{A}\,\vec{x} = \vec{\mathbf{0}}$ has only the zero solution.*
b. *The zero vector $\vec{\mathbf{0}}$ is a unique linear combination of the columns of $\mathbf{A}$.*
c. *The columns of the matrix $\mathbf{A}$ are linearly independent.*
d. *The reduced row echelon form of $\mathbf{A}$ has a leading one in each column.*
e. *The rank of the matrix $\mathbf{A}$ equals n.*

Proof. Let $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n \in \mathbb{K}^m$ denote the columns of the matrix $\mathbf{A}$.

$a \Rightarrow b$: We prove the contrapositive. When $\vec{\mathbf{0}} \in \mathbb{K}^n$ is a nonzero linear combination of the columns, there are scalars $v_1, v_2, \dots, v_n \in \mathbb{K}$, not all zero, such that $v_1\,\vec{a}_1 + v_2\,\vec{a}_2 + \cdots + v_n\,\vec{a}_n = \vec{\mathbf{0}}$. Hence, the vector $\vec{v} \in \mathbb{K}^n$ is a nonzero solution to the homogeneous linear system $\mathbf{A}\,\vec{x} = \vec{\mathbf{0}}$.

$b \Rightarrow c$: We again prove the contrapositive. When the columns are linearly dependent, there exists scalars $v_1, v_2, \dots, v_n \in \mathbb{K}$, not all zero, such that $v_1\,\vec{a}_1 + v_2\,\vec{a}_2 + \cdots + v_n\,\vec{a}_n = \vec{\mathbf{0}}$. Since we also have $0\,\vec{a}_1 + 0\,\vec{a}_2 + \cdots + 0\,\vec{a}_n = \vec{\mathbf{0}}$, we see that the zero vector $\vec{\mathbf{0}}$ is not a unique linear combination of the columns in $\mathbf{A}$.

$c \Rightarrow d$: Once again, we prove the contrapositive. Having a column in the reduced row echelon form of the matrix $\mathbf{A}$ that does not contain a leading one implies that $\operatorname{rank}(\mathbf{A}) < n$. Proposition 6.0.4 shows that the homogeneous linear system $\mathbf{A}\,\vec{x} = \vec{\mathbf{0}}$ has infinitely many solutions. Having a nonzero solution $\vec{v} \in \mathbb{K}^n$ to the homogeneous linear system $\mathbf{A}\,\vec{x} = \vec{\mathbf{0}}$ establishes that the vectors $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n \in \mathbb{K}^m$ are linearly dependent.

$d \Rightarrow a$: When the reduced row echelon form of the matrix $\mathbf{A}$ has a leading one in every column, we have $\operatorname{rank}(\mathbf{A}) = n$. Hence, Proposition 6.0.4 shows that the homogeneous linear system $\mathbf{A}\,\vec{x} = \vec{\mathbf{0}}$ has a unique solution.

$d \Leftrightarrow e$: From the definition of rank, we see that the rank of the matrix $\mathbf{A}$ equals n if and only if every column in the reduced row echelon form contains a leading one. $\square$

6.2.5 Problem. Determine whether the following sets of vectors are linearly dependent.

$$i.\ \left\{ \begin{bmatrix}1\\7\\6\end{bmatrix}, \begin{bmatrix}2\\0\\9\end{bmatrix}, \begin{bmatrix}3\\1\\5\end{bmatrix}, \begin{bmatrix}4\\1\\8\end{bmatrix} \right\} \quad ii.\ \left\{ \begin{bmatrix}2\\3\\5\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\8\end{bmatrix} \right\} \quad iii.\ \left\{ \begin{bmatrix}-2\\4\\6\\10\end{bmatrix}, \begin{bmatrix}3\\6\\9\\15\end{bmatrix} \right\}$$

Solution.

i. Let $\mathbf{A}$ be the (3×4)-matrix whose columns are the given vectors. The basic bounds on rank [4.2.3] show that $\operatorname{rank}(\mathbf{A}) \leqslant 3 < 4$, so the characterization of a unique solution proves that the given vectors are linearly dependent.

ii. Since $\vec{\mathbf{0}}$ belongs to the set, the vectors are linearly dependent.

iii. Although the second vector seems to be $\frac{3}{2}$ times the first vector, this relation only holds for the last three pairs of entries and fails for the first pair. Thus, neither of the vectors is a multiple of the other, so the vectors are linearly independent. □

Exercises

6.2.6 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. A linearly independent set has not linear relations.

ii. The zero vector can never belong to a linearly independent set.

iii. Every set with only one vector is linearly independent.

iv. The column vectors in a matrix are linearly independent if and only if the corresponding homogeneous linear system has infinitely many solutions.

v. If a matrix has more columns than rows, then the column vectors are linearly dependent.

vi. For a square matrix, its columns are linearly independent if and only if its rank equals the number of rows.

vii. For a square matrix, its columns are linearly independent if and only if they span the ambient coordinate space.

6.2.7 Problem. Prove that a set of vectors, which has more elements than the number of entries in any given vector, is linearly dependent.

7
Matrix Operations

The product of two matrices is more complicated than matrix arithmetic. In this chapter, we defined this binary operation, visualize it using directed graphs, and start to explore of invertibility.

7.0 Matrix Multiplication

How should we define matrix multiplication? When a matrix $\mathbf{A}$ multiplies a vector $\vec{v}$, it transforms $\vec{v}$ into the vector $\mathbf{A}\,\vec{v}$. If this vector is then multiplied by the matrix $\mathbf{B}$, the resulting vector is $\mathbf{B}(\mathbf{A}\,\vec{v})$. We want to represent the composite mapping $\vec{v} \mapsto \mathbf{B}(\mathbf{A}\,\vec{v})$ as multiplication by a single matrix $\mathbf{B}\,\mathbf{A}$, so that $\mathbf{B}(\mathbf{A}\,\vec{v}) = (\mathbf{B}\,\mathbf{A})\vec{v}$.

To motivate the formal definition, let $\mathbf{A}$ be an $(m \times n)$-matrix, let $\mathbf{B}$ an $(\ell \times m)$-matrix, and let $\vec{v} \in \mathbb{K}^n$. If $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2, \dots, \vec{\boldsymbol{a}}_n$ denote the columns of the matrix $\mathbf{A}$, then we have $\mathbf{A}\,\vec{x} = v_1\,\vec{\boldsymbol{a}}_1 + v_2\,\vec{\boldsymbol{a}}_2 + \cdots + v_n\,\vec{\boldsymbol{a}}_n \in \mathbb{K}^m$. The linearity of matrix multiplication [5.0.5] gives

$$\mathbf{B}(\mathbf{A}\,\vec{x}) = \mathbf{B}(v_1\,\vec{\boldsymbol{a}}_1) + \mathbf{B}(v_2\,\vec{\boldsymbol{a}}_2) + \cdots + \mathbf{B}(v_n\,\vec{\boldsymbol{a}}_n) = v_1(\mathbf{B}\vec{\boldsymbol{a}}_1) + v_2(\mathbf{B}\vec{\boldsymbol{a}}_2) + \cdots + v_n(\mathbf{B}\vec{\boldsymbol{a}}_n) \in \mathbb{K}^\ell .$$

In particular, we have $\mathbf{B}(\mathbf{A}\,\vec{x}) = \begin{bmatrix} \mathbf{B}\,\vec{\boldsymbol{a}}_1 & \mathbf{B}\,\vec{\boldsymbol{a}}_2 & \cdots & \mathbf{B}\,\vec{\boldsymbol{a}}_n \end{bmatrix} \vec{v}$.

7.0.0 Definition. Fix three positive integers ℓ, m, and n. Let $\mathbf{A}$ be an $(m \times n)$-matrix whose columns are the vectors $\vec{\boldsymbol{a}}_1, \vec{\boldsymbol{a}}_2, \dots, \vec{\boldsymbol{a}}_n \in \mathbb{K}^m$. For any $(\ell \times m)$-matrix $\mathbf{B}$, the ***matrix product*** $\mathbf{BA}$ is the $(\ell \times n)$-matrix whose columns are the vectors $\mathbf{B}\,\vec{\boldsymbol{a}}_1, \mathbf{B}\,\vec{\boldsymbol{a}}_2, \dots, \mathbf{B}\,\vec{\boldsymbol{a}}_n \in \mathbb{K}^\ell$.

We may compute the matrix product $\left[\begin{smallmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{smallmatrix}\right] \left[\begin{smallmatrix} 4 & 3 & 5 & 0 \\ 1 & -2 & 3 & -4 \end{smallmatrix}\right]$ in two slightly different ways. Since we have

$$\begin{bmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} 11 \\ 2 \\ -3 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 7 \\ 6 \end{bmatrix},$$
$$\begin{bmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} = 5 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} 19 \\ -1 \\ -9 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ -4 \end{bmatrix} = 0 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - 4 \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} -12 \\ 8 \\ 12 \end{bmatrix},$$

the matrix product is

$$\begin{aligned} \begin{bmatrix} 2 & 3 \\ 1 & -2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 4 & 3 & 5 & 0 \\ 1 & -2 & 3 & -4 \end{bmatrix} &= \begin{bmatrix} (2)(4)+(3)(1) & (2)(3)+(3)(-2) & (2)(5)+(3)(3) & (2)(0)+(3)(-4) \\ (1)(4)+(-2)(1) & (1)(3)+(-2)(-2) & (1)(5)+(-2)(3) & (1)(0)+(-2)(-4) \\ (0)(4)+(-3)(1) & (0)(3)+(-3)(-2) & (0)(5)+(-3)(3) & (0)(0)+(-3)(-4) \end{bmatrix} \\ &= \begin{bmatrix} 11 & 0 & 19 & -12 \\ 2 & 7 & -1 & 8 \\ -3 & 6 & -9 & 12 \end{bmatrix}. \end{aligned}$$

7.0.1 Remark. Matrix multiplication may be reinterpreted using the dot product. When the matrix product is defined, the (i,k)-entry of $\mathbf{B}\,\mathbf{A}$ is dot product of the i-th row in $\mathbf{B}$ with the k-th column in $\mathbf{A}$. For all $1 \leqslant i \leqslant \ell$, all $1 \leqslant j \leqslant m$, and all $1 \leqslant k \leqslant n$, let $b_{i,j} \in \mathbb{K}$ be the (i,j)-entry in $\mathbf{B}$ and let $a_{j,k} \in \mathbb{K}$ be the (j,k)-entry in $\mathbf{A}$. With this notation, the (i,k)-entry of the product $\mathbf{B}\,\mathbf{A}$ is

$$\sum_{j=1}^{m} b_{i,j}\, a_{j,k} = b_{i,1}\, a_{1,k} + b_{i,2}\, a_{2,k} + \cdots + b_{i,m}\, a_{m,k}\,.$$

7.0.2 Problem. Compute both products of the following matrices:

$$\mathbf{M} := \begin{bmatrix} 5 & 1 \\ 3 & -2 \\ 0 & 7 \end{bmatrix} \qquad\qquad \mathbf{N} := \begin{bmatrix} 2 & 0 & 1 \\ 4 & 3 & -1 \end{bmatrix}.$$

Solution. We have

$$\mathbf{M}\,\mathbf{N} = \begin{bmatrix} 5 & 1 \\ 3 & -2 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 4 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 14 & 3 & -4 \\ -2 & -6 & 6 \\ 28 & 21 & -7 \end{bmatrix}, \quad \mathbf{N}\,\mathbf{M} = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 3 & -1 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ 3 & -2 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 10 & 9 \\ 29 & -5 \end{bmatrix}.$$

These matrix products do not have the same size and no two entries are equal. $\square$

For matrix multiplication, the matrices giving rise to the identity operator have a simple description.

7.0.3 Definition. For any positive integer n, the ***identity matrix*** is the $(n \times n)$-matrix

$$\mathbf{I}_n := \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} \delta_{j,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The subscript on the square matrix $\mathbf{I}$ are frequently omitted when the number of rows or columns is clear from the context.

7.0.4 Proposition (Properties of matrix multiplication)**.** *Let* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *be matrices for which the indicated sums and products are defined. For any scalar* $c \in \mathbb{K}$, *we have the following properties:*

(identity)	$\mathbf{I}_m\,\mathbf{A} = \mathbf{A} = \mathbf{A}\,\mathbf{I}_n$	$(\mathbf{B}\,\mathbf{A})^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}}\,\mathbf{B}^{\mathsf{T}}$
(associativity)	$\mathbf{C}(\mathbf{B}\,\mathbf{A}) = (\mathbf{C}\,\mathbf{B})\mathbf{A}$	$c(\mathbf{B}\,\mathbf{A}) = (c\,\mathbf{B})\mathbf{A} = \mathbf{B}(c\,\mathbf{A})$
(distributivity)	$\mathbf{C}(\mathbf{B}+\mathbf{A}) = \mathbf{C}\,\mathbf{B} + \mathbf{C}\,\mathbf{A}$	$(\mathbf{C}+\mathbf{B})\mathbf{A} = \mathbf{C}\,\mathbf{A} + \mathbf{B}\,\mathbf{A}$

Proof. For all $1 \leqslant j \leqslant m$ and all $1 \leqslant k \leqslant n$, let $a_{j,k} \in \mathbb{K}$ denote the (j,k)-entry in the matrix $\mathbf{A}$ and let $\vec{a}_k \in \mathbb{K}^m$ denote the k-th column of $\mathbf{A}$. It follows that $\vec{a}_k = \begin{bmatrix} a_{1,k} & a_{2,k} & \cdots & a_{m,k} \end{bmatrix}^{\mathsf{T}}$. Since $\vec{e}_j \cdot \vec{a}_k = a_{j,k}$ and $\vec{a}_k \cdot \vec{e}_j = a_{k,j}$, we deduce that $\mathbf{I}_m\,\mathbf{A} = \mathbf{A} = \mathbf{A}\,\mathbf{I}_n$.

Similarly, for all $1 \leqslant i \leqslant \ell$ and all $1 \leqslant j \leqslant m$, let $b_{i,j} \in \mathbb{K}$ be the (i,j)-entry in the matrix $\mathbf{B}$. The definition of matrix multiplication implies that the (i,k)-entry in $\mathbf{B}\,\mathbf{A}$ is $\sum_{j=1}^{m} b_{i,j}\,a_{j,k}$ and the (k,i)-entry in $\mathbf{A}^\mathsf{T}\,\mathbf{B}^\mathsf{T}$ is $\sum_{j=1}^{m} a_{j,k}\,b_{i,j}$, so the commutativity of multiplication of scalars implies that $(\mathbf{B}\,\mathbf{A})^\mathsf{T} = \mathbf{A}^\mathsf{T}\,\mathbf{B}^\mathsf{T}$.

Finally, the general definition of matrix multiplication and the linearity of matrix multiplication [5.0.5] yield

$$\mathbf{C}(\mathbf{B}\,\mathbf{A}) = [\mathbf{C}(\mathbf{B}\,\vec{a}_1)\ \ \mathbf{C}(\mathbf{B}\,\vec{a}_2)\ \ \cdots\ \ \mathbf{C}(\mathbf{B}\,\vec{a}_n)] = [(\mathbf{C}\,\mathbf{B})\vec{a}_1\ \ (\mathbf{C}\,\mathbf{B})\vec{a}_2\ \ \cdots\ \ (\mathbf{C}\,\mathbf{B})\vec{a}_n] = (\mathbf{C}\,\mathbf{B})\mathbf{A}\,,$$

$$\begin{aligned} c(\mathbf{B}\,\mathbf{A}) &= [c(\mathbf{B}\,\vec{a}_1)\ \ c(\mathbf{B}\,\vec{a}_2)\ \ \cdots\ \ c(\mathbf{B}\,\vec{a}_n)] = [(c\,\mathbf{B})\vec{a}_1\ \ (c\,\mathbf{B})\vec{a}_2\ \ \cdots\ \ (c\,\mathbf{B})\vec{a}_n] = (c\,\mathbf{B})\mathbf{A} \\ &= [c(\mathbf{B}\,\vec{a}_1)\ \ c(\mathbf{B}\,\vec{a}_2)\ \ \cdots\ \ c(\mathbf{B}\,\vec{a}_n)] = [\mathbf{B}(c\,\vec{a}_1)\ \ \mathbf{B}(c\,\vec{a}_2)\ \ \cdots\ \ \mathbf{B}(c\,\vec{a}_n)] = \mathbf{B}(c\,\mathbf{A})\,, \end{aligned}$$

$$\begin{aligned} \mathbf{C}(\mathbf{B}+\mathbf{A}) &= \big[\mathbf{C}(\vec{b}_1+\vec{a}_1)\ \ \mathbf{C}(\vec{b}_2+\vec{a}_2)\ \ \cdots\ \ \mathbf{C}(\vec{b}_n+\vec{a}_n)\big] \\ &= \big[\mathbf{C}\,\vec{b}_1+\mathbf{C}\,\vec{a}_1\ \ \mathbf{C}\,\vec{b}_2+\mathbf{C}\,\vec{a}_2\ \ \cdots\ \ \mathbf{C}\,\vec{b}_n+\mathbf{C}\,\vec{a}_n\big] = \mathbf{C}\,\mathbf{B}+\mathbf{C}\,\mathbf{A}\,, \end{aligned}$$

$$\begin{aligned} (\mathbf{C}+\mathbf{B})\mathbf{A} &= [(\mathbf{C}+\mathbf{B})\vec{a}_1\ \ (\mathbf{C}+\mathbf{B})\vec{a}_2\ \ \cdots\ \ (\mathbf{C}+\mathbf{B})\vec{a}_n] \\ &= [\mathbf{C}\,\vec{a}_1+\mathbf{B}\,\vec{a}_1\ \ \mathbf{C}\,\vec{a}_2+\mathbf{B}\,\vec{a}_2\ \ \cdots\ \ \mathbf{C}\vec{a}_n+\mathbf{B}\,\vec{a}_n] = \mathbf{C}\,\mathbf{A}+\mathbf{B}\,\mathbf{A} \end{aligned}$$

which establishes the other four properties. □

7.0.5 Warning. There are some significant differences between matrix products and products of scalars.

- The product of two matrices depends on the order of the factors:

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

- The product of two nonzero matrices may equal zero:

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

- The cancellation property does not hold for matrix multiplication. The matrix equation $\mathbf{AB} = \mathbf{AC}$ does not imply $\mathbf{B} = \mathbf{C}$:

$$\begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix} = \begin{bmatrix}0&0\\0&0\end{bmatrix} = \begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}0&1\\0&0\end{bmatrix}, \text{ but } \begin{bmatrix}1&0\\0&0\end{bmatrix} \neq \begin{bmatrix}0&1\\0&0\end{bmatrix}.$$

7.0.6 Notation. For any square matrix $\mathbf{A}$ and any nonnegative integer k, the k-fold product of $\mathbf{A}$ is denoted by

$$\mathbf{A}^k := \underbrace{\mathbf{A}\,\mathbf{A}\cdots\mathbf{A}}_{k\text{ times}};$$

it is product of k copies of $\mathbf{A}$. The "empty" product is $\mathbf{A}^0 = \mathbf{I}$.

7.0.7 Problem. For $\mathbf{N} = \begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}$, compute $\mathbf{N}^k$ for all $k \in \mathbb{N}$.

Solution. We have $\mathbf{N}^0 = \mathbf{I}$, $\mathbf{N}^1 = \mathbf{N}$,

$$\mathbf{N}^2 = \begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix} = \begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix}, \quad \mathbf{N}^3 = \begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix} = \begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix},$$

and $\mathbf{N}^k = \mathbf{0}$ for all $k \geqslant 3$. □

Exercises

7.0.8 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. Matrix multiplication is defined for any two matrices.
ii. The product of two matrices is a matrix.
iii. The entry in a product of matrices equals the dot product of the corresponding row and column.
iv. The identity matrix can have any size.
v. Matrix multiplication is never commutative.

7.0.9 Problem. Let $\mathbf{A}$ be a square matrix. If the linear system $\mathbf{A}^2\,\vec{x} = \vec{\mathbf{0}}$ has infinitely many solutions, then prove that the linear system $\mathbf{A}\vec{x} = \vec{\mathbf{0}}$ also has infinitely many solutions.

7.0.10 Problem. The ***trace*** of an $(n \times n)$-matrix $\mathbf{A} := [a_{j,k}]$ is the sum of its diagonal entries:

$$\operatorname{tr}(\mathbf{A}) := a_{1,1} + a_{2,2} + \cdots + a_{n,n} = \sum_{j=1}^{n} a_{j,j}\,.$$

i. Prove that the trace is linear. In other words, show that, for any $(n \times n)$-matrices $\mathbf{A}$, $\mathbf{B}$ and any scalars $c, d \in \mathbb{K}$, we have $\operatorname{tr}(c\,\mathbf{A} + d\,\mathbf{B}) = c\operatorname{tr}(\mathbf{A}) + d\ \operatorname{tr}(\mathbf{B})$.
ii. If $\mathbf{A}$ and $\mathbf{B}$ are $(n \times n)$-matrices, then prove that $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$.
iii. For $n \geqslant 1$, show that the matrix equation $\mathbf{XY} - \mathbf{YX} = \mathbf{I}$ has no solutions for $(n \times n)$-matrices $\mathbf{X}$ and $\mathbf{Y}$.

7.1 Adjacency Matrices

HOW CAN WE VISUALIZE MATRIX MULTIPLICATION? A mathematical structure called a directed graph provides one answer.

7.1.0 Definition. A ***directed graph*** G is an order pair (V, E) where
- $V := \{1, 2, \ldots, n\}$ is a finite set whose elements are called ***vertices***,
- E is a set of ordered pairs of vertices called ***edges***.

The edge $(j, k) \in E$ is drawn as an arrow from the j-th vertex to the k-th vertex. The j-th vertex is the ***tail*** of the edge (j, k) and the k-th vertex is the ***head*** of the same edge.

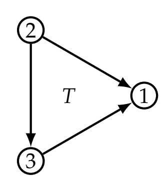

Figure 7.0: A small directed graph

The defining data of a directed graph can be encoded in a matrix.

7.1.1 Definition. For any directed graph G, the ***adjacency matrix*** $\mathbf{A}_G$ is a square matrix whose rows and columns correspond to the vertices of G. The (j, k)-entry in $\mathbf{A}_G$ is the number of edges in G from the k-th vertex to the j-th vertex.

For our two explicit examples, the adjacency matrice and some of their powers are:

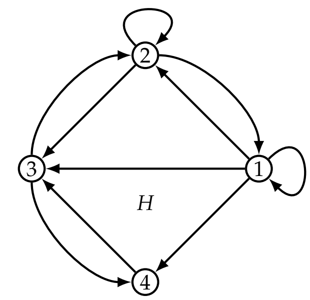

Figure 7.1: A larger directed graph

$$\mathbf{A}_T = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{A}_T^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{A}_T^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}_H = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{A}_H^2 = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 3 & 3 & 1 & 1 \\ 3 & 2 & 2 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix} \qquad \mathbf{A}_H^3 = \begin{bmatrix} 5 & 5 & 2 & 1 \\ 8 & 7 & 4 & 1 \\ 7 & 7 & 2 & 2 \\ 5 & 4 & 3 & 0 \end{bmatrix}$$

$$\mathbf{A}_H^4 = \begin{bmatrix} 13 & 12 & 6 & 2 \\ 10 & 19 & 8 & 4 \\ 18 & 16 & 9 & 2 \\ 12 & 12 & 4 & 3 \end{bmatrix} \qquad \mathbf{A}_H^5 = \begin{bmatrix} 33 & 31 & 14 & 6 \\ 51 & 47 & 23 & 8 \\ 45 & 43 & 18 & 9 \\ 31 & 28 & 15 & 4 \end{bmatrix} \qquad \mathbf{A}_H^6 = \begin{bmatrix} 84 & 78 & 37 & 14 \\ 129 & 121 & 55 & 23 \\ 115 & 106 & 52 & 18 \\ 78 & 74 & 32 & 15 \end{bmatrix}$$

7.1.2 Definition. A ***walk*** in the directed graph G is a sequence of edges $e_1, e_2, \ldots, e_\ell$ where the tail of e_{j+1} equals the head of e_j for all $1 \leqslant j < \ell$. The ***length*** of a walk is the number of edges in the sequence. Each vertex gives a distinct walk of length 0.

For example, the directed graph T has a unique path of length 2 from the second vertex to first vertex.

Multiplication of the adjacency matrix with itself beautifully enumerates all of the walks in a directed graph.

7.1.3 Proposition. *Let $\mathbf{A}_G$ be the adjacency matrix of a directed graph G. For any nonnegative integer ℓ, the (j,k)-entry of $\mathbf{A}_G^\ell$ is the number of walks with length ℓ from the k-th vertex to the j-th vertex.*

Inductive proof. When $\ell = 0$, we have $\mathbf{A}_G^0 = \mathbf{I}$ and the vertices are the walks of length 0, so the base case holds. Assume the proposition holds for some nonnegative integer ℓ and consider $\mathbf{A}_G^{\ell+1} = \mathbf{A}_G^\ell \, \mathbf{A}_G$. Let n be the number of vertices in G. For all $1 \leqslant j \leqslant n$ and all $1 \leqslant k \leqslant n$, let $a_{j,k}$ denote the (j,k)-entry in adjacency matrix $\mathbf{A}_G$. The definition of the adjacency matrix guarantees that $a_{j,k}$ equals the number of walks of length 1 in G from the k-th vertex to the j-th vertex. Similarly, if $b_{i,j}$ denotes the (i,j)-entry in $\mathbf{A}_G^\ell$ for all $1 \leqslant i \leqslant n$ and all $1 \leqslant j \leqslant n$, then the induction hypothesis implies that $b_{i,j}$ equals the number of walks of length ℓ in G from the j-th vertex to the k-th vertex. By concatenating walks at the j-th vertex, we see that there are exactly $b_{i,j}\, a_{j,k}$ walks of length $\ell + 1$ in G from the k-th vertex to the i-th vertex passing through the j-th vertex at second step. By summing over all possible choices for the j-th vertex, we see that there are $b_{i,1}\, a_{1,k} + b_{i,2}\, a_{2,k} + \cdots + b_{i,n}\, a_{n,k}$ walks of length $\ell + 1$ in G from the k-th vertex to the i-th vertex. From the dot-product reinterpretation of matrix multiplication [7.0.1], we recognize this sum as the (i,k)-entry in the product $\mathbf{A}_G^\ell \, \mathbf{A}_G = \mathbf{A}_G^{\ell+1}$. $\square$

To extend this ideas to non-square matrices, we focus on a special class of directed graphs.

7.1.4 Definition. A ***bipartite*** directed graph is a directed graph whose vertices can be divided into two disjoint sets such that the first contains all tails and the second contains all heads. In particular, no vertex is both a tail and a head.

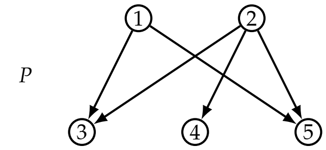

Figure 7.2: A bipartite directed graph

7.1.5 Definition. For a bipartite directed graph G, the ***biadjacency*** matrix $\mathbf{B}_G$ is a matrix whose columns correspond to the vertices that are the tail of some edge and whose rows correspond to the vertices that are the head of some edge. The (j,k)-entry of the biadjacency $\mathbf{B}_G$ is the number of edges in G from the k-th vertex to the j-th vertex.

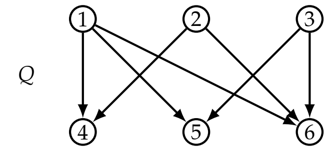

Figure 7.3: Another bipartite directed graph

Generalizing our insight for adjacency matrices, the product of biadjacency matrices also has an elegant combinatorial interpretation.

7.1.6 Proposition. *Let G and H be bipartite directed graphs such that the heads in G correspond to the tails in H. If $\mathbf{B}_G$ and $\mathbf{B}_H$ are the associated biadjacency matrices, then the (i,k)-entry of the product $\mathbf{B}_H\,\mathbf{B}_G$ is the number of walks from the k-th tail in G to the i-th head in H in the directed graph obtained by identifying the heads in G with the tails in H.*

The biadjacency matrices for our two bipartite directed graphs are

$$\mathbf{B}_P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$
$$\mathbf{B}_Q = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} .$$

Proof. Let ℓ be the number of heads in H, let m be the number of tails in H (or equivalently heads in G), and let n be the number of tails in G. For all $1 \leqslant j \leqslant m$ and all $1 \leqslant k \leqslant n$, let $a_{j,k}$ denote the (j,k)-entry in the adjacency matrix $\mathbf{B}_G$. The definition of the biadjacency matrix ensures that $a_{j,k}$ equals the number of walks of length 1 in G from the k-th vertex to the j-th vertex. Similarly, let $b_{i,j}$, for all $1 \leqslant i \leqslant \ell$ and all $1 \leqslant j \leqslant m$, denotes the (i,j)-entry in the adjacency matrix $\mathbf{B}_H$. Again, the definition of the biadjacency matrix ensures that $b_{i,j}$ equals the number of walks of length 1 in H from the i-th vertex to the j-th vertex. By concatenating walks at the j-th vertex, we see that there are $b_{i,j}\,a_{j,k}$ walks of length 2 in the merged graph from the k-th vertex the i-th vertex passing through the j-th vertex. Summing over all possible choices for the intermediate vertex, we see that there are $b_{i,1}\,a_{1,k} + b_{i,2}\,a_{2,k} + \cdots + b_{i,m}\,a_{m,k}$ walks of length 2 in the merged graph from the k-th vertex to the i-th vertex. From the dot-product reinterpretation of matrix multiplication [7.0.1], we recognize this sum as the (i,k)-entry in the product $\mathbf{B}_H\,\mathbf{B}_G$. □

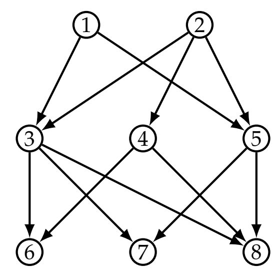

Figure 7.4: Merging two bipartite directed graphs

The product of biadjacency matrices is

$$\mathbf{B}_Q\,\mathbf{B}_P = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 3 \end{bmatrix}$$

7.1.7 Remark. To create an equivalence between directed graphs and matrices, we consider weighted graphs. A directed graph is ***weighted*** if each edge (k,j) has a weight $a_{j,k} \in \mathbb{K}$. In the weighted biadjacency matrix $\mathbf{B}_G$, the (j,k)-entry equals the weight $a_{j,k} \in \mathbb{K}$. Every matrix is the weighted biadjacency matrix of some weighted bipartite directed graph. The ***weight*** of a walk to be the product of the weights on its edges; a walk of length 0 has weight 1. Each entry in a product of matrices may be viewed as the sum of weighted walks between the appropriate vertices in a weighted directed graph.