APPENDIX D

MATRIX ALGEBRA REVIEW

D.1 BASIC OPERATIONS

A *matrix* is a rectangular array of elements. Rectangles are arranged in rows and columns. The general form of an $m \times n$ matrix (m rows and n columns) is

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

The element a_{ij} is the entry in the ith row and jth column of $\boldsymbol{A}$. An $n \times n$ matrix is said to be *square*.

A *row vector* is a $1 \times n$ matrix. A *column vector* is an $n \times 1$ matrix. Matrices are denoted with bold uppercase letters. Vectors are denoted with bold lowercase letters. The ith component of the vector $\boldsymbol{x}$ is denoted x_i. The Euclidean space $\mathbb{R}^n$ consists of all n-element vectors of real numbers.

Given n-element vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the *dot product*, or *inner product*, of $\boldsymbol{x}$ and $\boldsymbol{y}$ is the number

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

Introduction to Stochastic Processes with R, First Edition. Robert P. Dobrow.

A real number is called a *scalar*. If s is a scalar and $\boldsymbol{A}$ a matrix, then *scalar multiplication* is defined as the product $s\boldsymbol{A}$, which is the matrix obtained by multiplying each of the elements of $\boldsymbol{A}$ by s. For example,

$$(-3)\begin{pmatrix} 4 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -12 & 0 \\ 3 & -3 \end{pmatrix}.$$

Matrix addition is the operation of adding two matrices of the same dimension. Corresponding elements are added. For example,

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

Linear Combination

A *linear combination* of vectors $\boldsymbol{x_1}, \ldots, \boldsymbol{x_k}$ is a vector of the form

$$s_1\boldsymbol{x_1} + \cdots + s_k\boldsymbol{x_k},$$

where $s_1, \ldots, s_k$ are scalars. The s_i are called the *coefficients* of the linear combination.

Observe that $\begin{pmatrix} 8 & -1 & -3 \end{pmatrix}$ is a linear combination of $\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 2 & -1 & 1 \end{pmatrix}$, and $\begin{pmatrix} 3 & 3 & 3 \end{pmatrix}$ since

$$\begin{pmatrix} 8 & -1 & -3 \end{pmatrix} = (2)\begin{pmatrix} 1 & 1 & 0 \end{pmatrix} + (3)\begin{pmatrix} 2 & -1 & 1 \end{pmatrix} + (0)\begin{pmatrix} 3 & 3 & 3 \end{pmatrix}.$$

The coefficients of this linear combination are 2, 3, and 0.

We define the matrix–vector product $\boldsymbol{Ax}$ of an $m \times n$ matrix $\boldsymbol{A}$ and an $n \times 1$ column vector $\boldsymbol{x}$. Write the columns of $\boldsymbol{A}$ as $\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}$. Then, $\boldsymbol{Ax}$ is defined as the $m \times 1$ column vector, which is the linear combination of the columns of $\boldsymbol{A}$ whose coefficients are the components of $\boldsymbol{x}$. That is,

$$\boldsymbol{Ax} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n} \\ | & | & \cdots & | \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\boldsymbol{a_1} + x_2\boldsymbol{a_2} + \cdots + x_n\boldsymbol{a_n}.$$

The ith component of $\boldsymbol{Ax}$ is

$$(\boldsymbol{Ax})_i = \sum_{j=1}^{n} a_{ij}x_j, \text{ for } i = 1, \ldots, m.$$

Equivalently, the ith component of $\boldsymbol{Ax}$ is the dot product of the ith row of $\boldsymbol{A}$ and the vector $\boldsymbol{x}$. For example,

$$\begin{pmatrix} 2 & 1 & 1 \\ 3 & -1 & 2 \\ 1 & 0 & -4 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 2(1) + 1(1) + 1(-2) \\ 3(1) + (-1)(1) + 2(-2) \\ 1(1) + 0(1) + (-4)(-2) \\ 4(1) + 2(1) + 1(-2) \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 9 \\ 4 \end{pmatrix}$$

$$= (1)\begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix} + (1)\begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix} + (-2)\begin{pmatrix} 1 \\ 2 \\ -4 \\ 1 \end{pmatrix}.$$

D.2 LINEAR SYSTEM

A *linear system* of m equations in n unknowns is a collection of linear equations of the form

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

Such a system can be written succinctly in matrix form as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where

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

Linear systems can have no solutions, infinitely many solutions, or exactly one solution. For instance, the system

$$\begin{array}{l} 2x_1 + x_2 = 5 \\ 2x_1 + x_2 = 4 \end{array}$$

has no solutions. The system

$$\begin{array}{l} 2x_1 + x_2 = 5 \\ 4x_1 + 2x_2 = 10 \end{array}$$

has infinitely many solutions of the form

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a \\ 5 - 2a \end{pmatrix}, \quad \text{for all real } a.$$

And the system

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

has the unique solution $\boldsymbol{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$.

D.3 MATRIX MULTIPLICATION

If $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is an $n \times p$ matrix, then the matrix product $\boldsymbol{AB}$ is defined as the $m \times p$ matrix whose ith column is the matrix–vector product of $\boldsymbol{A}$ and the ith column of $\boldsymbol{B}$. Writing

$$\boldsymbol{B} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_p} \\ | & | & \cdots & | \end{pmatrix}$$

gives

$$\boldsymbol{AB} = \boldsymbol{A} \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_p} \\ | & | & \cdots & | \end{pmatrix} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{Ab_1} & \boldsymbol{Ab_2} & \cdots & \boldsymbol{Ab_p} \\ | & | & \cdots & | \end{pmatrix}.$$

The ijth element of $\boldsymbol{AB}$ is

$$(\boldsymbol{AB})_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Equivalently, the ijth element of $\boldsymbol{AB}$ is the dot product of the ith row of $\boldsymbol{A}$ and the jth column of $\boldsymbol{B}$. For example,

$$\begin{pmatrix} 0 & 1 & 4 \\ 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 1 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 12 \\ 4 & 3 \end{pmatrix}.$$

Matrix multiplication is not commutative That is, $\boldsymbol{AB}$ does not necessarily equal $\boldsymbol{BA}$.

D.4 DIAGONAL, IDENTITY MATRIX, POLYNOMIALS

Given an $n \times n$ matrix $\boldsymbol{A}$, the entries $a_{11}, \ldots, a_{nn}$ are called the diagonal elements of $\boldsymbol{A}$. An $n \times n$ matrix $\boldsymbol{A}$ is a *diagonal matrix* if $a_{ij} = 0$, for all $i \neq j$.

The $n \times n$ *identity matrix*, denoted $\boldsymbol{I}_n$, is the diagonal matrix all of whose diagonal elements are 1. The columns of the $n \times n$ identity matrix are called the *standard basis vectors of* $\mathbb{R}^n$, denoted $\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}$. That is, $\boldsymbol{e_k}$ is the n-element column vector of all 0s except for a 1 in the kth position.

If $\boldsymbol{A}$ is an $n \times n$ matrix, then $\boldsymbol{AI}_n = \boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$.

If $\boldsymbol{A}$ is a square matrix, then $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}^2$ is also a square matrix. Similarly, $\boldsymbol{A}^k$ is well-defined for all integer k. It follows that if $p(x)$ is a polynomial function and $\boldsymbol{A}$ is a square matrix, then $p(\boldsymbol{A})$ is well-defined. For instance, let $p(x) = x^3 - 5x + 6$ and $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Then,

$$p(\boldsymbol{A}) = \boldsymbol{A}^3 - 5\boldsymbol{A} + 6\boldsymbol{I} = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} - (5)\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + (6)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ -3 & 7 \end{pmatrix}.$$

D.5 TRANSPOSE

Given an $m \times n$ matrix $\boldsymbol{A}$, the *transpose* $\boldsymbol{A}^T$ is the $n \times m$ matrix whose ijth element is the jith element of $\boldsymbol{A}$.

A matrix $\boldsymbol{A}$ is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}^T$. That is, $a_{ij} = a_{ji}$ for all i, j. A symmetric matrix is necessarily square.

D.6 INVERTIBILITY

A square matrix $\boldsymbol{A}$ is *invertible* if there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. The matrix $\boldsymbol{B}$ is denoted $\boldsymbol{A}^{-1}$ and is called *the inverse of* $\boldsymbol{A}$.

Since $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, it follows that the ith column of $\boldsymbol{A}^{-1}$ is the solution of the linear system $\boldsymbol{Ax} = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the ith standard basis vector.

The solution of a general linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is unique if and only if $\boldsymbol{A}$ is invertible. In that case, the solution is $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$.

Properties of Inverse, Transpose

1. $(\boldsymbol{A}^T)^T = \boldsymbol{A}$
2. $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
3. $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$
4. $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
5. $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

D.7 BLOCK MATRICES

It is sometimes convenient to partition a matrix $\boldsymbol{A}$ into smaller *blocks*, such as

$$\boldsymbol{A} = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{B} & \boldsymbol{C} \\ \hline \boldsymbol{D} & \boldsymbol{E} \end{array}\right),$$

where

$$\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 13 & 14 & 15 \end{pmatrix}, \text{ and } \quad \boldsymbol{E} = \begin{pmatrix} 16 \end{pmatrix}.$$

Matrix operations on block matrices can be carried out by treating the blocks as matrix elements. Thus,

$$\begin{aligned} \boldsymbol{A}^2 &= \left(\begin{array}{c|c} \boldsymbol{B} & \boldsymbol{C} \\ \hline \boldsymbol{D} & \boldsymbol{E} \end{array}\right) \left(\begin{array}{c|c} \boldsymbol{B} & \boldsymbol{C} \\ \hline \boldsymbol{D} & \boldsymbol{E} \end{array}\right) \\ &= \left(\begin{array}{c|c} \boldsymbol{B}^2 + \boldsymbol{CD} & \boldsymbol{BC} + \boldsymbol{CE} \\ \hline \boldsymbol{DB} + \boldsymbol{ED} & \boldsymbol{DC} + \boldsymbol{E}^2 \end{array}\right) \\ &= \left(\begin{array}{ccc|c} 90 & 100 & 110 & 120 \\ 202 & 228 & 254 & 280 \\ 314 & 356 & 398 & 440 \\ \hline 426 & 484 & 542 & 600 \end{array}\right). \end{aligned}$$

D.8 LINEAR INDEPENDENCE AND SPAN

Given a set of vectors, if at least one vector can be written as a linear combination of the others, then the vectors are called *linearly dependent*. If none of the vectors in the set can be written as a linear combination of the other vectors, then the vectors are called *linearly independent*.

The vectors $\left\{ \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ -3 \\ 7 \end{pmatrix} \right\}$ are linearly dependent, as

$$\begin{pmatrix} 4 \\ -3 \\ 7 \end{pmatrix} = (2) \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} + (-1) \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

The vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ are linearly independent.

The *span* of a set of vectors is the set of all linear combinations of those vectors.

The span of $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ is the set of all vectors of the form

$$a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}, \quad \text{for scalars } a \text{ and } b.$$

Geometrically, this set is the x–y plane in $\mathbb{R}^3$.

D.9 BASIS

A *basis* for $\mathbb{R}^n$ is a set of n vectors $\{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ in $\mathbb{R}^n$, which are linearly independent and span $\mathbb{R}^n$. The fact that $\{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ span $\mathbb{R}^n$ means that every vector in $\mathbb{R}^n$ can be written as a linear combination of the $\boldsymbol{b_i}$. Together with linear independence we obtain the following result.

Theorem D.1. *If $\{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ is a basis for $\mathbb{R}^n$, then every vector in $\mathbb{R}^n$ can be written uniquely as a linear combination of the $\boldsymbol{b_i}$.*

To obtain this unique representation for a given set of vectors $\{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$, let $\boldsymbol{B}$ be the square matrix obtained by making the $\boldsymbol{b_i}$ the columns of $\boldsymbol{B}$. That is,

$$\boldsymbol{B} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_n} \\ | & | & \cdots & | \end{pmatrix}.$$

The matrix $\boldsymbol{B}$ is invertible. If $\boldsymbol{x} = s_1\boldsymbol{b_1} + \cdots + s_n\boldsymbol{b_n}$ for some choice of $s_1, \dots, s_n$, then

$$\boldsymbol{x} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_n} \\ | & | & \cdots & | \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \boldsymbol{B} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$$

and thus

$$\begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \boldsymbol{B}^{-1}\boldsymbol{x}.$$

D.10 VECTOR LENGTH

The *length* (also magnitude or norm) of a vector $\boldsymbol{x}$ is defined as

$$\| \boldsymbol{x} \| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}.$$

A vector $\boldsymbol{x}$ has unit length if $\| \boldsymbol{x} \| = 1$. For any nonzero vector $\boldsymbol{v}$, the vector $\left(\frac{1}{\|\boldsymbol{v}\|}\right)\boldsymbol{v}$ has unit length, since

$$\left\| \left(\frac{1}{\| \boldsymbol{v} \|}\right) \boldsymbol{v} \right\| = \left(\frac{1}{\| \boldsymbol{v} \|}\right) \| \boldsymbol{v} \| = 1.$$

A most important inequality in linear algebra is the Cauchy–Schwarz inequality, which says that for any n-element vectors $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \| \boldsymbol{x} \| \| \boldsymbol{y} \| .$$

For example, letting $\boldsymbol{x} = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}$, the inequality yields

$$|y_1 + \cdots + y_n| \le \sqrt{n}\sqrt{y_1^2 + \cdots + y_n^2}.$$

Equivalently,

$$(y_1 + \cdots + y_n)^2 \le n(y_1^2 + \cdots + y_n^2).$$

D.11 ORTHOGONALITY

Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be *orthogonal* if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. Geometrically, orthogonal vectors are perpendicular.

A set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is *orthonormal* if the vectors have unit length and are pairwise orthogonal. That is,

$$\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

A square matrix whose columns are orthonormal is called an *orthogonal matrix*. If $\boldsymbol{U}$ is an orthogonal matrix, then $\boldsymbol{U}$ is invertible and $\boldsymbol{U}^{-1} = \boldsymbol{U}^T$. The latter follows from the definition of orthonormal vectors since the ijth element of $\boldsymbol{U}^T\boldsymbol{U}$ is the dot product of the ith column of $\boldsymbol{U}$ and the jth column of $\boldsymbol{U}$.

The matrix

$$\boldsymbol{U} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{2/3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}$$

is an orthogonal matrix as

$$\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{2/3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -\sqrt{2/3} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

D.12 EIGENVALUE, EIGENVECTOR

Let $\boldsymbol{A}$ be a square matrix. If there exists a scalar λ and nonzero (column) vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, we say that λ is an *eigenvalue* of $\boldsymbol{A}$ with corresponding *eigenvector* $\boldsymbol{x}$.

For example, let $\boldsymbol{A} = \begin{pmatrix} 3 & 0 \\ 1 & -4 \end{pmatrix}$. Observe that

$$\begin{pmatrix} 3 & 0 \\ 1 & -4 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \end{pmatrix} = (-4)\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

which shows that $\lambda = -4$ is an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Further observe that

$$\begin{pmatrix} 3 & 0 \\ 1 & -4 \end{pmatrix} \begin{pmatrix} 7 \\ 1 \end{pmatrix} = \begin{pmatrix} 21 \\ 3 \end{pmatrix} = (3) \begin{pmatrix} 7 \\ 1 \end{pmatrix},$$

which shows that $\lambda = 3$ is an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector $\begin{pmatrix} 7 \\ 1 \end{pmatrix}$.

One can show that $\lambda = -4$ and $\lambda = 3$ are the only eigenvalues of $\boldsymbol{A}$.

D.13 DIAGONALIZATION

A square matrix $\boldsymbol{A}$ is *diagonalizable* if there exists an invertible matrix $\boldsymbol{S}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{A} = \boldsymbol{SDS}^{-1}$. If $\boldsymbol{A}$ is diagonalizable then the entries of $\boldsymbol{D}$ are the eigenvalues of $\boldsymbol{A}$ and the columns of $\boldsymbol{S}$ are corresponding eigenvectors.

For example, let $\boldsymbol{A} = \begin{pmatrix} 6 & -3 & 7 \\ 6 & -3 & 6 \\ 0 & 0 & -1 \end{pmatrix}$. The eigenvalues of $\boldsymbol{A}$ are $\lambda = 3, -1, 0$, with respective eigenvectors $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \right\}$.

Let $\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}$, with $\boldsymbol{S}^{-1} = \begin{pmatrix} 2 & -1 & 2 \\ 0 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix}$. This gives

$$\begin{aligned} \boldsymbol{SDS}^{-1} &= \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 & 2 \\ 0 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 6 & -3 & 7 \\ 6 & -3 & 6 \\ 0 & 0 & -1 \end{pmatrix} = \boldsymbol{A}. \end{aligned}$$

That is, $\boldsymbol{A}$ is diagonalizable.

A sufficient condition for an $n \times n$ matrix to be diagonalizable is that there exists n distinct eigenvalues. The following theorem gives a necessary and sufficient condition for diagonalizability.

Theorem D.2. *An $n \times n$ matrix $\boldsymbol{A}$ is diagonalizable if and only if there exists a basis for $\mathbb{R}^n$ consisting of eigenvectors of $\boldsymbol{A}$.*

One advantage of diagonalizability is that it simplifies matrix products. If $\boldsymbol{A}$ is diagonalizable, then

$$\boldsymbol{A}^k = (\boldsymbol{SDS}^{-1})^k = \boldsymbol{SD}^k\boldsymbol{S}^{-1}.$$

In the previous example, for $k \geq 1$,

$$\begin{aligned}\boldsymbol{A}^k = \boldsymbol{SD}^k\boldsymbol{S}^{-1} &= \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 3^k & 0 & 0 \\ 0 & (-1)^k & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & -1 & 2 \\ 0 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} (2)3^k & -3^k & (2)3^k - (-1)^k \\ (2)3^k & -3^k & (2)3^k \\ 0 & 0 & (-1)^k \end{pmatrix}.\end{aligned}$$

A square matrix $\boldsymbol{A}$ is *orthogonally diagonalizable* if there exists an orthogonal matrix $\boldsymbol{U}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{A} = \boldsymbol{UDU}^T$.

Theorem D.3 *(Spectral Theorem).* A matrix is orthogonally diagonalizable if and only if it is symmetric.