Introduction to Linear Algebra

Jason R. Wilson

Contents

CHAPTER 1

Matrix Operations

We use $\mathbb{R}^{m\times n}$ to denote the set of all matrices with m rows and n columns whose entries are in $\mathbb{R}$. For a matrix $A \in \mathbb{R}^{m\times n}$, the entry in row i and column j is denoted a_{ij}. Let $A, B \in \mathbb{R}^{m\times n}$ and $c \in \mathbb{R}$ where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$

We define

$$A + B = \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

$$cA = \begin{bmatrix} c\,a_{11} & \cdots & c\,a_{1n} \\ \vdots & \ddots & \vdots \\ c\,a_{m1} & \cdots & c\,a_{mn} \end{bmatrix}$$

We define $\mathbb{R}^n$ to be the set of $n \times 1$ matrices $\mathbb{R}^{n\times 1}$ and we use $\boldsymbol{x}_i$ to denote the i^{th} component of the column vector $\boldsymbol{x} \in \mathbb{R}^n$. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \qquad c\boldsymbol{x} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}$$

1.1. Matrix Multiplication

Definition 1.1

Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$. The matrix product $AB \in \mathbb{R}^{m\times p}$ is defined using the formula

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} \tag{1.1}$$

Note that AB is defined only if the number of columns of A equals the number of rows of B.

Consider the matrices

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

We calculate

$$AB = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad BA = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Note that $AB \neq BA$. Thus, matrix multiplication is not commutative.

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Suppose A has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ where

$$\boldsymbol{a}_i = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix}$$

Using equation (1.1) we calculate

$$\begin{aligned} A\boldsymbol{x} &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \begin{bmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{bmatrix} + \cdots + \begin{bmatrix} a_{1n}x_n \\ a_{2n}x_n \\ \vdots \\ a_{mn}x_n \end{bmatrix} \\ &= x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n \end{aligned}$$

and hence

$$\boxed{A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n} \tag{1.2}$$

Equation (1.2) says that $A\boldsymbol{x}$ is a linear combination of the column vectors of A where the coefficients are determined by the components of x.

Let $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $c, d \in \mathbb{R}$. Suppose A has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Using equation (1.2) we calculate

$$\begin{aligned} A(c\boldsymbol{x} + d\boldsymbol{y}) &= (cx_1 + dy_1)\boldsymbol{a}_1 + (cx_2 + dy_2)\boldsymbol{a}_2 + \cdots + (cx_n + dy_n)\boldsymbol{a}_n \\ &= (cx_1\boldsymbol{a}_1 + cx_2\boldsymbol{a}_2 + \cdots + cx_n\boldsymbol{a}_n) + (dy_1\boldsymbol{a}_1 + dy_2\boldsymbol{a}_2 + \cdots + dy_n\boldsymbol{a}_n) \\ &= cA\boldsymbol{x} + dA\boldsymbol{y} \end{aligned}$$

and hence

$$\boxed{A(c\boldsymbol{x} + d\boldsymbol{y}) = cA\boldsymbol{x} + dA\boldsymbol{y}} \tag{1.3}$$

We say that matrix vector multiplication is a linear operation.

Let $A \in \mathbb{R}^{3\times 2}$ and suppose that $A\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}3\\0\\1\end{bmatrix}$ and $A\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}5\\1\\0\end{bmatrix}$. We would like to calculate $A\begin{bmatrix}2\\1\end{bmatrix}$ without explicity finding the matrix A. We note that

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

and hence by equation (1.3) we have

$$\begin{aligned} A\begin{bmatrix}2\\1\end{bmatrix} &= A\left(3\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix}\right) \\ &= 3A\begin{bmatrix}1\\1\end{bmatrix} - A\begin{bmatrix}1\\2\end{bmatrix} \\ &= 3\begin{bmatrix}3\\0\\1\end{bmatrix} - \begin{bmatrix}5\\1\\0\end{bmatrix} \\ &= \begin{bmatrix}4\\-1\\3\end{bmatrix} \end{aligned}$$

Exercise 1.1

Let $A \in \mathbb{R}^{m\times n}$, $c_1, c_2, \ldots, c_k \in \mathbb{R}$, and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$. Use induction on $k \geq 2$ to show that

$$A(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = c_1A\boldsymbol{v}_1 + c_2A\boldsymbol{v}_2 + \cdots + c_kA\boldsymbol{v}_k \tag{1.4}$$

Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$. Suppose B has column vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_p$. Using equation (1.1) we calculate

$$\begin{aligned}(AB)_{ij} &= \sum_{k=1}^{n} a_{ik}b_{kj} \\ &= \sum_{k=1}^{n} a_{ik}(\boldsymbol{b}_j)_k \\ &= (A\boldsymbol{b}_j)_i \\ &= \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_p \end{bmatrix}_{ij}\end{aligned}$$

and hence

$$\boxed{AB = \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_p \end{bmatrix}} \tag{1.5}$$

Next, let $\boldsymbol{x} \in \mathbb{R}^p$. Using equations (1.2), (1.4), and (1.5) we calculate

$$\begin{aligned}(AB)\boldsymbol{x} &= \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_p \end{bmatrix} \boldsymbol{x} \\ &= x_1 A\boldsymbol{b}_1 + x_2 A\boldsymbol{b}_2 + \cdots + x_p A\boldsymbol{b}_p \\ &= A(x_1\boldsymbol{b}_1 + x_2\boldsymbol{b}_2 + \cdots + x_p\boldsymbol{b}_p) \\ &= A(B\boldsymbol{x})\end{aligned}$$

and hence

$$\boxed{(AB)x = A(Bx)} \tag{1.6}$$

Finally, let $C \in \mathbb{R}^{p\times q}$. Suppose C has column vectors $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_q$. Using equations (1.5) and (1.6) we calculate

$$\begin{aligned}(AB)C &= \begin{bmatrix} (AB)\boldsymbol{c}_1 & (AB)\boldsymbol{c}_2 & \cdots & (AB)\boldsymbol{c}_q \end{bmatrix} \\ &= \begin{bmatrix} A(B\boldsymbol{c}_1) & A(B\boldsymbol{c}_2) & \cdots & A(B\boldsymbol{c}_q) \end{bmatrix} \\ &= A\begin{bmatrix} B\boldsymbol{c}_1 & B\boldsymbol{c}_2 & \cdots & B\boldsymbol{c}_q \end{bmatrix} \\ &= A(BC)\end{aligned}$$

We have shown that

$$\boxed{(AB)C = A(BC)} \tag{1.7}$$

and thus matrix multiplication is associative.

1.2. Matrix Inverse and Transpose

The identity matrix, $I_n \in \mathbb{R}^{n\times n}$, has column vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \in \mathbb{R}^n$ where

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots \quad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Exercise 1.2

Let $A, B \in \mathbb{R}^{m\times n}$. Show that if $A\boldsymbol{e}_i = B\boldsymbol{e}_i$ for all i in $1, \ldots, n$, then $A = B$.

Definition 1.2

The matrix $A \in \mathbb{R}^{n\times n}$ is called invertible if there exists a $B \in \mathbb{R}^{n\times n}$ such that

$$AB = BA = I_n$$

Let $A \in \mathbb{R}^{n\times n}$. Suppose that

$$AB = BA = I_n, \qquad AC = CA = I_n$$

for some $B, C \in \mathbb{R}^{n\times n}$. Using equation (1.7) we calculate

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

We have shown that If A is invertible, then its inverse is unique. We use A^{-1} to denote the unique inverse of an invertible matrix A.

Lemma 1.2.1

Let $A \in \mathbb{R}^{n\times n}$. If A is invertible, then its inverse is unique.

Suppose that $A \in \mathbb{R}^{n\times n}$ is invertible. Since

$$A^{-1}A = I_n, \qquad AA^{-1} = I_n$$

we have that A^{-1} is invertible and

$$\boxed{(A^{-1})^{-1} = A} \tag{1.8}$$

Let $A, B \in \mathbb{R}^{n\times n}$ where A and B are invertible. Using equation (1.7) we calculate

$$(AB)(B^{-1}A^{-1}) = A(B(B^{-1}A^{-1})) = A((BB^{-1})A^{-1}) = A(I_nA^{-1}) = AA^{-1} = I_n$$

$$(B^{-1}A^{-1})(AB) = ((B^{-1}A^{-1})A)B = (B^{-1}(A^{-1}A))B = (B^{-1}I_n)B = B^{-1}B = I_n$$

We have shown that if A and B are invertible then AB is invertible and

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}} \tag{1.9}$$

Exercise 1.3

Let $A_1, A_2, \ldots A_k \in \mathbb{R}^{n\times n}$ and suppose that A_i is invertible for all $i \in 1, \ldots k$. Use induction on $k \geq 2$ to show that $A_1A_2 \cdots A_k$ is invertible and that

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1} \tag{1.10}$$

Definition 1.3

Let $A \in \mathbb{R}^{m\times n}$. We define $A^T \in \mathbb{R}^{n\times m}$ by the formula

$$(A^T)_{ij} = a_{ji} \tag{1.11}$$

Note that $(A^T)^T = A$. If $A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 5 & 7 \end{bmatrix}$, then $A^T = \begin{bmatrix} 2 & -1 \\ 1 & 5 \\ 0 & 7 \end{bmatrix}$.

Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$. We calculate

$$\begin{aligned}(AB)^T_{ij} &= (AB)_{ji} \\ &= \sum_{k=1}^{n} a_{jk} b_{ki} \\ &= \sum_{k=1}^{n} b_{ki} a_{jk} \\ &= \sum_{k=1}^{n} (B^T)_{ik}(A^T)_{kj} \\ &= (B^T A^T)_{ij}\end{aligned}$$

We have shown that

$$\boxed{(AB)^T = B^T A^T} \tag{1.12}$$

Exercise 1.4

Suppose $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, and $C \in \mathbb{R}^{p\times q}$. Show that

$$(ABC)^T = C^T B^T A^T \tag{1.13}$$

Exercise 1.5

Let $A \in \mathbb{R}^{n\times n}$ and suppose that A is invertible. Show that A^T is invertible and that

$$(A^T)^{-1} = (A^{-1})^T \tag{1.14}$$

Definition 1.4

A matrix $A \in \mathbb{R}^{n\times n}$ is called symmetric if $A^T = A$.

Exercise 1.6

Let $A \in \mathbb{R}^{m\times n}$. Show that $A^T A$ and AA^T are symmetric matrices.

Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times n}$. Suppose that A has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and B has column vectors vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$. We first note that for any k satisfying $1 \le k \le n$ we have

$$\boldsymbol{a}_k = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}, \quad \boldsymbol{b}_k^T = \begin{bmatrix} b_{1k} & b_{2k} & \cdots & b_{pk} \end{bmatrix}, \quad (\boldsymbol{a}_k \boldsymbol{b}_k^T)_{ij} = a_{ik} b_{jk}$$

Using the definition of matrix multiplication we note that

$$\begin{aligned}(AB^T)_{ij} &= a_{i1}b_{j1} + a_{i2}b_{j2} + \cdots + a_{in}b_{jn} \\ &= (\boldsymbol{a}_1\boldsymbol{b}_1^T)_{ij} + (\boldsymbol{a}_2\boldsymbol{b}_2^T)_{ij} + \cdots + (\boldsymbol{a}_n\boldsymbol{b}_n^T)_{ij} \\ &= (\boldsymbol{a}_1\boldsymbol{b}_1^T + \boldsymbol{a}_2\boldsymbol{b}_2^T + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^T)_{ij}\end{aligned}$$

We have shown that

$$\boxed{AB^T = \boldsymbol{a}_1\boldsymbol{b}_1^T + \boldsymbol{a}_2\boldsymbol{b}_2^T + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^T} \tag{1.15}$$

1.3. Elementary Row Operations

Definition 1.5

Let $A \in \mathbb{R}^{m\times n}$. The following three operations on the rows of A are called elementary row operations.

(1) Interchange any two rows of A.
(2) Multiply any row of A be a nonzero scalar.
(3) Add any scalar multiple of a row of A to another row of A.

The result of applying a single elementary row operation to an identity matrix is called an elementary matrix.

Let $A \in \mathbb{R}^{3\times 4}$ be given by

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

and let $E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. Note that E_1 is I_3 with rows 1 and 3 interchanged and is thus an elementary matrix. To find E_1^{-1} apply the reverse of the elementary row operation to I_3. To reverse interchanging rows 1 and 3, we interchange rows 1 and 3 again. It is easy to check that E_1 is invertible and $E_1^{-1} = E_1$. We calculate

$$E_1 A = \begin{bmatrix} 2 & 2 & -2 & -2 \\ 2 & 3 & -2 & -3 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

and note that $E_1 A$ is the matrix A with rows 1 and 3 interchanged.

Next let $B = E_1 A$ and $E_2 = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Note that E_2 is I_3 with row 1 mutliplied by 1/2 and is thus an elementary matrix. To find E_2^{-1} apply the reverse of the elementary row operation to I_3. To reverse multiplying row 1 by 1/2, we multiply row 1 by 2. It is easy to check that E_2 is invertible and that $E_2^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is also an elementary matrix. We calculate

$$E_2 B = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 3 & -2 & -3 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

and note that $E_2 B$ is the matrix B with row 1 multiplied by 1/2.

Next let $C = E_2 B$ and $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Note that E_3 is I_3 with -2 times row 1 added to row 2 and is thus an elementary matrix. To find E_3^{-1} apply the reverse of the elementary row operation to I_3. To reverse adding -2 times row 1 to row 2, we add 2 times row 1 to row 2. It is easy to check that E_3 is invertible and that $E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is also an elementary matrix. We calculate

$$E_3 C = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

and note that $E_3 C$ is the matrix C with -2 times row 1 added to row 2.

Let $D = E_3C$ and note that $D = PA$ where $P = E_3E_2E_1$ is an invertible matrix with $P^{-1} = E_1^{-1}E_2^{-1}E_3^{-1}$.

In general, suppose k elementary operations are performed on the matrix A resulting in the matrix C. Let E_i be the elementary matrix that encodes the i^{th} elementary row operation. Then we have $C = PA$ where $P = E_kE_{k-1}\cdots E_1$. Since elementary row operations can be reversed, all elementary matrices are invertible. Thus P is invertible with $P^{-1} = E_1^{-1}E_2^{-1}\cdots E_k^{-1}$. We have outlined the proof of the following theorem.

Theorem 1.1

Let $A \in \mathbb{R}^{m\times n}$. Let C be the matrix resulting from applying any number of elementary row operations to A. Then there exists an invertible matrix $P \in \mathbb{R}^{m\times m}$ such that $C = PA$.

Exercise 1.7

Consider the matrices

$$A = \begin{bmatrix} 1 & -2 & 0 & 2 & -3 \\ 2 & -4 & 2 & 0 & 8 \\ 1 & -2 & 3 & -3 & 16 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & -2 & 0 & 2 & -3 \\ 0 & 0 & 1 & -2 & 7 \\ 0 & 0 & 3 & -5 & 19 \end{bmatrix}$$

(1) Find an invertible matrix $P \in \mathbb{R}^{3\times 3}$ that is the product of elementary matrices such that
$$C = PA$$
(2) Write P^{-1} as a product of elementary matrices.
(3) Explicitly calculate P^{-1} and $P^{-1}C$. Verify that $A = P^{-1}C$.

CHAPTER 2

Solving Linear Systems

The set of equations

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

is called a *system of linear equations* or *linear system*. The linear system can be written as in matrix form as $A\boldsymbol{x} = \boldsymbol{b}$ where

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

A linear system that has a solution is called *consistent*. A linear system with no solution is called *inconsistent*. A linear system is called *homogeneous* if $\boldsymbol{b} = 0$ and *nonhomogeneous* otherwise. A linear system is called *underdetermined* if there are more variables than equations and *overdetermined* if there are more equations than variables.

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Let C be the matrix resulting from applying any number of elementary row operations to A and $\boldsymbol{d} \in \mathbb{R}^m$ be the column vector resulting from applying the same elementary row operations to $\boldsymbol{b}$. By theorem 1.1 there exists an invertible matrix $P \in \mathbb{R}^{m\times m}$ such that $C = PA$ and $\boldsymbol{d} = P\boldsymbol{b}$. Let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary. If $A\boldsymbol{x} = \boldsymbol{b}$, then

$$C\boldsymbol{x} = (PA)\boldsymbol{x} = P(A\boldsymbol{x}) = P\boldsymbol{b} = \boldsymbol{d}$$

If $C\boldsymbol{x} = \boldsymbol{d}$, then

$$A\boldsymbol{x} = (P^{-1}C)\boldsymbol{x} = P^{-1}(C\boldsymbol{x}) = P^{-1}\boldsymbol{d} = \boldsymbol{b}$$

We summarize these results in the following theorem.

Theorem 2.1

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Let C be the matrix resulting from applying any number of elementary row operations to A and $\boldsymbol{d} \in \mathbb{R}^m$ be the column vector resulting from applying the *same* elementary row operations to $\boldsymbol{b}$. For all $\boldsymbol{x} \in \mathbb{R}^n$ we have that $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $C\boldsymbol{x} = \boldsymbol{d}$.

2.1. Gauss-Jordan Elimination

Suppose that

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

The *augmented matrix* for the system $A\boldsymbol{x} = \boldsymbol{b}$ is the matrix

$$A|\boldsymbol{b} = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

Algorithm 2.1

Gauss-Jordan elimination uses elementary row operations to reduce a matrix (or augmented matrix) to reduced row echelon form. The algorithm consists of two passes.

(1) In the forward pass, the matrix is transformed into a matrix in which the first nonzero entry of each row is 1, and it occurs in a column to the right of the first nonzero entry of each preceding row.

(2) In the backward pass, the matrix is transformed into a matrix in which the first nonzero entry of each row is the only nonzero entry of its column.

Consider the linear system $A\boldsymbol{x} = \boldsymbol{0}$ where

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

We use Gauss-Jordan elimination to reduce A to reduced row echelon form. We first interchange the first row and the second row

$$\begin{bmatrix} 1 & -4 & -1 & 1 \\ 0 & -1 & 1 & -4 \\ -1 & 4 & -2 & 5 \end{bmatrix}$$

Adding the first row to the third row

$$\begin{bmatrix} 1 & -4 & -1 & 1 \\ 0 & -1 & 1 & -4 \\ 0 & 0 & -3 & 6 \end{bmatrix}$$

Multiplying the second row by -1

$$\begin{bmatrix} 1 & -4 & -1 & 1 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & -3 & 6 \end{bmatrix}$$

Multiplying the third row by $-1/3$

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

completes the first pass of Gauss-Jordan elimination. Next, we add the third row to the second row and add the third row to the first row

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

Finally, adding 4 times the second row to the first row

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

completes the second pass of Gauss-Jordan elimination and the matrix C is in reduced row echelon form. By theorem 2.1, the linear systems $A\boldsymbol{x} = \boldsymbol{0}$ and $C\boldsymbol{x} = \boldsymbol{0}$ have the same solutions. We can write the linear system $C\boldsymbol{x} = \mathbf{0}$ as

$$\begin{aligned} x_1 + 7x_4 &= 0 \\ x_2 + 2x_4 &= 0 \\ x_3 - 2x_4 &= 0 \end{aligned}$$

Solving for the leftmost variables gives

$$x_1 = -7x_4 \tag{2.1}$$

$$x_2 = -2x_4 \tag{2.2}$$

$$x_3 = 2x_4 \tag{2.3}$$

If we set $x_4 = 1$ we get the nonzero solution

$$\boldsymbol{x} = \begin{bmatrix} -7 \\ -2 \\ 2 \\ 1 \end{bmatrix}$$

Note that for *any* choice for x_4 we can use equations (2.1), (2.2), and (2.3) to *determine* the remaining variables x_1, x_2, and x_3. We call the variable x_4 a *free variable.* Setting the *free variable* x_4 to t we can write the general solution to the linear system $A\boldsymbol{x} = \mathbf{0}$ as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -7t \\ -2t \\ 2t \\ t \end{bmatrix} = t \begin{bmatrix} -7 \\ -2 \\ 2 \\ 1 \end{bmatrix}$$

Consider the linear system

$$\begin{bmatrix} 1 & -2 & 0 & 2 & -3 \\ 2 & -4 & 2 & 0 & 8 \\ 1 & -2 & 3 & -3 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 8 \end{bmatrix} \tag{2.4}$$

We apply Gauss-Jordan elimination to reduce the augmented matrix

$$\left[\begin{array}{ccccc|c} 1 & -2 & 0 & 2 & -3 & 2 \\ 2 & -4 & 2 & 0 & 8 & 6 \\ 1 & -2 & 3 & -3 & 16 & 8 \end{array}\right]$$

to reduced row echelon form. Adding -2 times the first row to the second row and -1 times the first row to the third row gives

$$\left[\begin{array}{ccccc|c} 1 & -2 & 0 & 2 & -3 & 2 \\ 0 & 0 & 2 & -4 & 14 & 2 \\ 0 & 0 & 3 & -5 & 19 & 6 \end{array}\right]$$

Multiplying the second row by $-1/2$

$$\left[\begin{array}{ccccc|c} 1 & -2 & 0 & 2 & -3 & 2 \\ 0 & 0 & 1 & -2 & 7 & 1 \\ 0 & 0 & 3 & -5 & 19 & 6 \end{array}\right]$$

Adding -3 times the second row to the third row

$$\left[\begin{array}{ccccc|c} 1 & -2 & 0 & 2 & -3 & 2 \\ 0 & 0 & 1 & -2 & 7 & 1 \\ 0 & 0 & 0 & 1 & -2 & 3 \end{array}\right]$$

completes the first pass of Gauss-Jordan elimination. Adding 2 times the third row to the second row and -2 times the third row to the first row

$$\left[\begin{array}{ccccc|c} 1 & -2 & 0 & 0 & 1 & -4 \\ 0 & 0 & 1 & 0 & 3 & 7 \\ 0 & 0 & 0 & 1 & -2 & 3 \end{array}\right] \tag{2.5}$$

completes the second pass of Gauss-Jordan elimination and the augmented matrix in equation (2.5) is in reduced row echelon form. By theorem 2.1 the solutions to the corresponding linear system

$$\begin{bmatrix} 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -4 \\ 7 \\ 3 \end{bmatrix} \tag{2.6}$$

are the same as the solutions to the linear system in equation (2.4). The linear system in equation (2.6) can be written as

$$\begin{aligned} x_1 - 2x_2 + x_5 &= -4 \\ x_3 + 3x_5 &= 7 \\ x_4 - 2x_5 &= 3 \end{aligned}$$

Solving for the leftmost variables in each equation gives

$$x_1 = 2x_2 - x_5 - 4 \tag{2.7}$$

$$x_3 = -3x_5 + 7 \tag{2.8}$$

$$x_4 = 2x_5 + 3 \tag{2.9}$$

If we set $x_2 = x_5 = 0$ we get the solution to equation (2.4) given by

$$\boldsymbol{x} = \begin{bmatrix} -4 \\ 0 \\ 7 \\ 3 \\ 0 \end{bmatrix}$$

Note that for *any* assignment of scalars to the variables x_2 and x_5 we can use equations (2.7), (2.8), and (2.9) to *determine* the remaining variables x_1, x_3, and x_4. We call the variables x_2 and x_5 *free variables.* Thinking of the free variables as *parameters* we set $t_1 = x_2$ and $t_2 = x_5$. The general solution to equation (2.4) can thus be written as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2t_1 - t_2 - 4 \\ t_1 \\ -3t_2 + 7 \\ 2t_2 + 3 \\ t_2 \end{bmatrix} = \begin{bmatrix} -4 \\ 0 \\ 7 \\ 3 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -1 \\ 0 \\ -3 \\ 2 \\ 1 \end{bmatrix}$$

2.2. An Inconsistent Linear System

Consider the linear system

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 3 & -1 & 4 \\ 1 & 4 & -3 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 6 \end{bmatrix} \tag{2.10}$$

We apply Gauss-Jordan elimination to the augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 5 \\ 2 & 3 & -1 & 4 & 8 \\ 1 & 4 & -3 & -3 & 6 \end{array}\right]$$

Adding -2 times the first row to the second row gives

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 5 \\ 0 & -1 & 1 & 2 & -2 \\ 1 & 4 & -3 & -3 & 6 \end{array}\right]$$

Adding -1 times the first row to third row gives

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 5 \\ 0 & -1 & 1 & 2 & -2 \\ 0 & 2 & -2 & -4 & 1 \end{array}\right]$$

Multiplying the second row by -1 gives

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 5 \\ 0 & 1 & -1 & -2 & 2 \\ 0 & 2 & -2 & -4 & 1 \end{array}\right]$$

Adding -2 times the second row to the third row gives

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

By theorem 2.1 the solutions to the linear system

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -3 \end{bmatrix} \tag{2.11}$$

are the same as the solutions to the linear system (2.10). Since it is impossible to choose scalars x_1, x_2, x_3, x_4 such that

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = -3$$

the linear system (2.11) has no solutions and hence the linear system (2.10) also has no solutions. We say that the linear system (2.10) is *inconsistent*.

Exercise 2.1

For each of the following linear systems:

- If necessary, write the system in the form $Ax = b$.
- Use Gauss-Jordan elimination to reduce A or $A|b$ to reduced row echelon form.
- Find the general solution of the linear system.

(1)

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

(2)

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 - x_2 + x_3 - x_4 &= 0 \end{aligned}$$

(3)

$$\begin{bmatrix} 0 & 2 & 0 & 2 \\ 1 & 2 & -1 & 3 \\ 2 & 4 & -1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$$

(4)

$$\begin{bmatrix} 1 & -1 & 2 & 3 & 1 \\ 2 & -3 & 6 & 9 & 4 \\ 3 & -1 & 2 & 4 & 1 \\ 7 & -2 & 4 & 8 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ 2 \\ 6 \end{bmatrix}$$

Exercise 2.2

Does the following linear system have a solution? Explain.

$$\begin{aligned} x_1 + x_2 - 3x_3 + x_4 &= 1 \\ x_1 + x_2 + x_3 - x_4 &= 2 \\ x_1 + x_2 - x_3 &= 0 \end{aligned}$$

2.3. Calculating Matrix Inverses

Let $A \in \mathbb{R}^{n \times n}$ and suppose that A is invertible. Consider the problem of finding $A^{-1} \in \mathbb{R}^{n \times n}$ with column vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$ such that $AA^{-1} = I_n$. The equation

$$\begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix} = I_n = AA^{-1} = A \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix} = \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_n \end{bmatrix}$$

shows that we must have $A\boldsymbol{b}_i = \boldsymbol{e}_i$ for all i in $1, \ldots, n$. We can find $\boldsymbol{b}_i$ by row reducing the augmented matrix $A|\boldsymbol{e}_i$ to reduced row echelon form which will output $I_n|\boldsymbol{b}_i$. We can find all n columns of A^{-1} simultaneously by row reducing the matrix $A|I_n$ to reduced row echelon form which will output $I_n|A^{-1}$.

Consider the matrix

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

We attempt to row reduce the augmented matrix $A|I_3$ to $I_3|A^{-1}$. We start with

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

Adding the first row to the second row and adding -1 times the first row to the third row gives

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

Multiplying the second row by 1/3 gives

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

Adding 2 times the second row to the third row gives

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

Multiplying the third row by 1/2 gives

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1 & -1/6 & 1/3 & 1/2 \end{array}\right]$$

Adding -1 times the third row to the second row and adding -1 times the third row to the first row gives

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 7/6 & -1/3 & -1/2 \\ 0 & 1 & 0 & 1/2 & 0 & -1/2 \\ 0 & 0 & 1 & -1/6 & 1/3 & 1/2 \end{array}\right]$$

Adding -2 times the second row to the first row gives

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/6 & -1/3 & 1/2 \\ 0 & 1 & 0 & 1/2 & 0 & -1/2 \\ 0 & 0 & 1 & -1/6 & 1/3 & 1/2 \end{array}\right]$$

We have shown that if A is invertible we must have

$$A^{-1} = \begin{bmatrix} 1/6 & -1/3 & 1/2 \\ 1/2 & 0 & -1/2 \\ -1/6 & 1/3 & 1/2 \end{bmatrix}$$

To verify our answer we calculate

$$AA^{-1} = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/6 & -1/3 & 1/2 \\ 1/2 & 0 & -1/2 \\ -1/6 & 1/3 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

$$A^{-1}A = \begin{bmatrix} 1/6 & -1/3 & 1/2 \\ 1/2 & 0 & -1/2 \\ -1/6 & 1/3 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

Next, we consider the matrix

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

We attempt to row reduce the augmented matrix $A|I_3$ to $I_3|A^{-1}$. We start with

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -1 & 1 & 0 & 0 \\ -4 & 3 & 0 & 0 & 1 & 0 \\ 2 & 0 & 6 & 0 & 0 & 1 \end{array}\right]$$

Adding -2 times the first row to the second row and 4 times the first row to the third row gives

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & -4 & 4 & 1 & 0 \\ 0 & 2 & 8 & -2 & 0 & 1 \end{array}\right]$$

Multiplying the second row by -1 gives

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 4 & -4 & -1 & 0 \\ 0 & 2 & 8 & -2 & 0 & 1 \end{array}\right]$$

Adding -2 times the second row to the third row gives

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

We now see that $A|\boldsymbol{e}_1$ row reduces to

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

Hence $A\boldsymbol{x} = \boldsymbol{e}_1$ has no solution and A is not invertible since if A was invertible the vector $A^{-1}\boldsymbol{e}_1$ would be a solution to $A\boldsymbol{x} = \boldsymbol{e}_1$.

2.4. Polynomial Interpolation

Consider the problem of finding a polynomial

$$f(t) = x_1 + x_2 t + x_3 t^2$$

that satisfies $f(1) = 8$, $f(2) = 5$, and $f(3) = -4$. The constraints yield the linear system

$$x_1 + x_2 + x_3 = 8$$

$$x_1 + 2x_2 + 4x_3 = 5$$

$$x_1 + 3x_2 + 9x_3 = -4$$

which can be written in matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ -4 \end{bmatrix} \tag{2.12}$$

We apply Gauss-Jordan elimination to the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 8 \\ 1 & 2 & 4 & 5 \\ 1 & 3 & 9 & -4 \end{array}\right]$$

Adding -1 times the first row to the second row and -1 times the first row to the third row gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 8 \\ 0 & 1 & 3 & -3 \\ 0 & 2 & 8 & -12 \end{array}\right]$$

Adding -2 times the second row to the third row gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 8 \\ 0 & 1 & 3 & -3 \\ 0 & 0 & 2 & -6 \end{array}\right]$$

Multiplying the third row by 1/2 gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 8 \\ 0 & 1 & 3 & -3 \\ 0 & 0 & 1 & -3 \end{array}\right]$$

Adding -3 times the third row to the second row and -1 times the third row to the first row gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 11 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & -3 \end{array}\right]$$

Adding -1 times the second row to the first row gives

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & -3 \end{array}\right]$$

By inspection we get the *unique* solution $\boldsymbol{x} = \begin{bmatrix} 5 \\ 6 \\ -3 \end{bmatrix}$ with corresponding interpolating polynomial $f(t) = 5 + 6t - 3t^2$. It is easy to check that $f(1) = 8$, $f(2) = 5$, and $f(3) = -4$.

Exercise 2.3

If possible, find a polynomial of the form

$$f(t) = x_1 + x_2 t + x_3 t^2 + x_4 t^3$$

such that $f(-2) = 3$, $f(-1) = -6$, $f(1) = 0$, and $f(3) = -2$. If you found such a polynomial satisfying the constraints, is it unique?

CHAPTER 3

Vector Spaces

A vector space is a set of objects called *vectors* together with an operation that adds two vectors and an operation that multiplies a scalar times a vector. These two operations must satisfy certain *vector space axioms*. This axiomatic definition enable us to prove generic results that apply to all vector spaces.

3.1. Definition and Examples

Definition 3.1

Definition: A vector space V (over $\mathbb{R}$) is a set of objects V called *vectors* with an addition and scalar multiplication such that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and $a, b \in \mathbb{R}$ the following axioms are satisfied.

- (Closure) $a\boldsymbol{x} + b\boldsymbol{y}$ in V
- (VS 1) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$
- (VS 2) $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$
- (VS 3) There exists $\mathbf{0} \in V$ such that $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$ for all $\boldsymbol{x} \in V$.
- (VS 4) For all $\boldsymbol{x} \in V$ there exists $\boldsymbol{y} \in V$ such that $\boldsymbol{x} + \boldsymbol{y} = \mathbf{0}$.
- (VS 5) $1\boldsymbol{x} = \boldsymbol{x}$
- (VS 6) $(ab)\boldsymbol{x} = a(b\boldsymbol{x})$
- (VS 7) $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}$
- (VS 8) $(a + b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}$

The set of column vectors $V = \mathbb{R}^n$ with the normal addition and scalar multiplication operations is a vector space. We verify axioms (VS 3), (VS 4), (VS 7) and leave the remainder of the proof as an exercise.

(1) (VS 3) Let $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ be the vector in $\mathbb{R}^n$ with all components equal to 0. Note that for any $\boldsymbol{x} \in \mathbb{R}^n$ we have that

$$\boldsymbol{x} + \mathbf{0} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + 0 \\ x_2 + 0 \\ \vdots \\ x_n + 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{x}$$

(2) (VS 4) Let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary. Define $\boldsymbol{y} = \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$ and note that

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix} = \begin{bmatrix} x_1 + (-x_1) \\ x_2 + (-x_2) \\ \vdots \\ x_n + (-x_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

(3) (VS 7) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$. We calculate

$$a(\boldsymbol{x}+\boldsymbol{y}) = a\begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} = \begin{bmatrix} a(x_1+y_1) \\ a(x_2+y_2) \\ \vdots \\ a(x_n+y_n) \end{bmatrix} = \begin{bmatrix} ax_1+ay_1 \\ ax_2+ay_2 \\ \vdots \\ ax_n+ay_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} ay_1 \\ ay_2 \\ \vdots \\ ay_n \end{bmatrix} = a\boldsymbol{x} + a\boldsymbol{y}$$

Exercise 3.1

Prove that the set of matrices $V = \mathbb{R}^{m\times n}$ with the normal addition and scalar multiplication operations is a vector space.

Let $V = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$ and define the addition and scalar mutliplication to be the usual operations for $\mathbb{R}^3$. Then V is not a vector space since $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is in V but $2\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ is not in V.

Let $V = \mathbb{R}^2$ and for $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2$ and $c \in \mathbb{R}$, define

$$\boldsymbol{x}+\boldsymbol{y} = \begin{bmatrix} x_1+y_1 \\ 2x_2+y_2 \end{bmatrix}, \qquad c\boldsymbol{x} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$$

Since

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

axiom (VS 1) does not hold and the set V with the given operations is not a vector space.

Exercise 3.2

Let $V = \mathbb{R}^2$. For $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in V$ and $c \in \mathbb{R}$, define

$$\boldsymbol{x}+\boldsymbol{y} = \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \end{bmatrix}, \qquad c\boldsymbol{x} = \begin{bmatrix} cx_1 \\ -cx_2 \end{bmatrix}$$

Is V with the given operations a vector space? Explain.

Let $\mathcal{F}(S,\mathbb{R})$ denote the set of all functions from a set S to $\mathbb{R}$. Two functions $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{F}(S,\mathbb{R})$ are equal if $\boldsymbol{f}(t) = \boldsymbol{g}(t)$ for all $t \in S$. For $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{F}(S,\mathbb{R})$ define $\boldsymbol{f}+\boldsymbol{g}$ and $c\boldsymbol{f}$ to be the functions satisfying

$$(\boldsymbol{f}+\boldsymbol{g})(t) = \boldsymbol{f}(t)+\boldsymbol{g}(t), \qquad (c\boldsymbol{f})(t) = c\boldsymbol{f}(t)$$

Then, the set $\mathcal{F}(S,\mathbb{R})$ with the given operations is a vector space. We verify axioms (VS 3), (VS 4), (VS 8) and leave the remainder of the proof as an exercise.

(1) (VS 3) Let $\boldsymbol{0}$ be the function whose value is 0 for all t in S. Let $\boldsymbol{f} \in \mathcal{F}(S,\mathbb{R})$ be arbitrary. For any $t \in S$ we have

$$(\boldsymbol{f}+\boldsymbol{0})(t) = \boldsymbol{f}(t)+\boldsymbol{0}(t) = \boldsymbol{f}(t)+0 = \boldsymbol{f}(t)$$

and thus

$$\boldsymbol{f}+\boldsymbol{0} = \boldsymbol{f}$$

(2) (VS 4) Let $\boldsymbol{f} \in \mathcal{F}(S,\mathbb{R})$ be arbitrary. Let $\boldsymbol{g} \in \mathcal{F}(S,\mathbb{R})$ be the function whose values are given by $\boldsymbol{g}(t) = -\boldsymbol{f}(t)$. For any $t \in S$ we have

$$(\boldsymbol{f}+\boldsymbol{g})(t) = \boldsymbol{f}(t)+\boldsymbol{g}(t) = \boldsymbol{f}(t)+(-\boldsymbol{f}(t)) = 0 = \boldsymbol{0}(t)$$

and thus

$$\boldsymbol{f}+\boldsymbol{g} = \boldsymbol{0}$$

(3) (VS 8) Let $\boldsymbol{f} \in \mathcal{F}(S, \mathbb{R})$ and $a, b \in \mathbb{R}$ be arbitrary. For any $t \in S$ we have

$$((a+b)\boldsymbol{f})(t) = (a+b)\boldsymbol{f}(t) = a\boldsymbol{f}(t) + b\boldsymbol{f}(t) = (a\boldsymbol{f})(t) + (b\boldsymbol{f})(t) = (a\boldsymbol{f} + b\boldsymbol{f})(t)$$

and thus

$$(a+b)\boldsymbol{f} = a\boldsymbol{f} + b\boldsymbol{f}$$

3.2. Basic Results

Let V be a vector space. By (VS 3), there exists $\mathbf{0} \in V$ such that for all $\boldsymbol{x} \in V$

$$\boldsymbol{x} + \mathbf{0} = \boldsymbol{x} \tag{3.1}$$

Suppose that for some potentially different vector $\mathbf{0}' \in V$ we also have that for all $\boldsymbol{x} \in V$

$$\boldsymbol{x} + \mathbf{0}' = \boldsymbol{x} \tag{3.2}$$

We note that

$$\begin{aligned} \mathbf{0} &= \mathbf{0} + \mathbf{0}' && \text{Equation (3.2)} \\ &= \mathbf{0}' + \mathbf{0} && \text{Axiom (VS 2)} \\ &= \mathbf{0}' && \text{Equation (3.1)} \end{aligned}$$

We have justified the following result.

Lemma 3.2.1

Let V be a vector space. Then there is a *unique* vector $\mathbf{0} \in V$ such that

$$\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$$

for all $\boldsymbol{x} \in V$.

Let V be a vector space and suppose that

$$\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y} + \boldsymbol{z} \tag{3.3}$$

for vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$. By (VS 4) there exists a vector $\boldsymbol{w} \in V$ such that

$$\boldsymbol{z} + \boldsymbol{w} = \mathbf{0} \tag{3.4}$$

We note that

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x} + \mathbf{0} && \text{Axiom (VS 3)} \\ &= \boldsymbol{x} + (\boldsymbol{z} + \boldsymbol{w}) && \text{Equation (3.4)} \\ &= (\boldsymbol{x} + \boldsymbol{z}) + \boldsymbol{w} && \text{Axiom (VS 2)} \\ &= (\boldsymbol{y} + \boldsymbol{z}) + \boldsymbol{w} && \text{Equation (3.3)} \\ &= \boldsymbol{y} + (\boldsymbol{z} + \boldsymbol{w}) && \text{Axiom (VS 2)} \\ &= \boldsymbol{y} + \mathbf{0} && \text{Equation (3.4)} \\ &= \boldsymbol{y} && \text{Axiom (VS 3)} \end{aligned}$$

We have justified the following result.

Lemma 3.2.2

Let V be a vector space. If $\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y} + \boldsymbol{z}$ where $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, then $\boldsymbol{x} = \boldsymbol{y}$.

Let V be a vector space and suppose that

$$\boldsymbol{z} + \boldsymbol{x} = \boldsymbol{z} + \boldsymbol{y} \tag{3.5}$$

We note that

$$\begin{aligned} \boldsymbol{x} + \boldsymbol{z} &= \boldsymbol{z} + \boldsymbol{x} && \text{Axiom (VS 2)} \\ &= \boldsymbol{z} + \boldsymbol{y} && \text{Equation (3.5)} \\ &= \boldsymbol{y} + \boldsymbol{z} && \text{Axiom (VS 2)} \end{aligned}$$

and thus $\boldsymbol{x} = \boldsymbol{y}$ by lemma 3.2.2. We have justified the following theorem.

Theorem 3.1: Cancellation Law

Let V be a vector space. If $\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y} + \boldsymbol{z}$ or $\boldsymbol{z} + \boldsymbol{x} = \boldsymbol{z} + \boldsymbol{y}$ where $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, then $\boldsymbol{x} = \boldsymbol{y}$.

Let V be a vector space and let $\boldsymbol{x} \in V$ be arbitrary. By (VS 4), there exists $\boldsymbol{y} \in V$ such that

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0} \tag{3.6}$$

Suppose that for some potentially different vector $\boldsymbol{z} \in V$ we also have that

$$\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{0} \tag{3.7}$$

Then $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$ and by theorem 3.1 we have $\boldsymbol{y} = \boldsymbol{z}$. We have justified the following result.

Corollary 3.2.1

Let V be a vector space and let $\boldsymbol{x} \in V$ be arbitrary. Then there exists a unique $\boldsymbol{y} \in V$ such that

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0}$$

We denote the unique additive inverse of $\boldsymbol{x}$ by $-\boldsymbol{x}$ and define $\boldsymbol{x} - \boldsymbol{y}$ to mean $\boldsymbol{x} + (-\boldsymbol{y})$.
Let V be a vector space and let $\boldsymbol{x} \in V$ be arbitrary. We note that

$$\begin{aligned} 0\boldsymbol{x} + \boldsymbol{0} &= 0\boldsymbol{x} && \text{Axiom (VS 3)} \\ &= (0+0)\boldsymbol{x} \\ &= 0\boldsymbol{x} + 0\boldsymbol{x} && \text{Axiom (VS 8)} \end{aligned}$$

and thus $0\boldsymbol{x} = \boldsymbol{0}$ by theorem 3.1. We have justified the following result

Corollary 3.2.2

Let V be a vector space. Then for all $\boldsymbol{x} \in V$ we have that

$$0\boldsymbol{x} = \boldsymbol{0}$$

Exercise 3.3

Let V be a vector space. Then for all $a \in \mathbb{R}$ we have that

$$a\boldsymbol{0} = \boldsymbol{0}$$

Let V be a vector space and let $\boldsymbol{x} \in V$ be arbitrary. We note that

$$\begin{aligned} \boldsymbol{x} + (-1)\boldsymbol{x} &= 1\boldsymbol{x} + (-1)\boldsymbol{x} && \text{Axiom (VS 5)} \\ &= (1 + (-1))\boldsymbol{x} && \text{Axiom (VS 8)} \\ &= 0\boldsymbol{x} \\ &= \boldsymbol{0} && \text{Corollary 3.2.2} \end{aligned}$$

and thus $(-1)\boldsymbol{x} = -\boldsymbol{x}$ by corollary 3.2.1. We have justified the following result.

Corollary 3.2.3

Let V be a vector space and let $\boldsymbol{x} \in V$ be arbitrary. Then

$$(-1)\boldsymbol{x} = -\boldsymbol{x}$$

Exercise 3.4

Let V be a vector space. Prove that

$$(-c)\boldsymbol{x} = -(c\boldsymbol{x}) = c(-x)$$

for all $c \in \mathbb{R}$ and all $\boldsymbol{x} \in V$.

3.3. Subspaces

Definition 3.2

A subset W of a vector space V is called a *subspace* of V if W is a vector space with the operations of addition and scalar multiplication defined on V.

Note that V is a subspace of any vector space V. Suppose W is a subspace of a vector space V. By axiom (VS3), W must contain an additive identity $\mathbf{0}_W$. Since $W \subseteq V$ we have that $\mathbf{0}_W \in V$ as well. Let $\mathbf{0}_V$ be the unique zero vector for V. By corollary 3.2.2 we have that $0(\mathbf{0}_W) = \mathbf{0}_V$. But then axiom (Closure) applied to W implies that $\mathbf{0}_V \in W$. Let $\boldsymbol{x} \in W$. Then $\boldsymbol{x} \in V$ and by axiom (VS3) applied to V we have $\boldsymbol{x} + \mathbf{0}_V = \boldsymbol{x}$. Thus $\mathbf{0}_V$ works as an additive identity for W and by lemma 3.2.1 we must have $\mathbf{0}_W = \mathbf{0}_V$. We have justified the following result.

Lemma 3.3.1

Suppose W is a subspace of a vector space V. Then $\mathbf{0}_V \in W$. Furthermore, $\mathbf{0}_V$ is the unique zero vector for W.

The next theorem gives us a way to check if a subset W of a vector space V is a subspace of V.

Theorem 3.2

A subset W of a vector space V is a subspace of V if and only if

(1) $\mathbf{0}_V \in W$
(2) $a\boldsymbol{x} + b\boldsymbol{y} \in W$ for all $\boldsymbol{x}, \boldsymbol{y} \in W$ and $a, b \in \mathbb{R}$.

Proof. Let W be a subset of a vector space V. First suppose that W is a subspace of V. By lemma 3.3.1 we have that $\mathbf{0}_V \in W$ and thus (1) holds. Since W is a vector space, axiom (Closure) implies that (2) holds. Next suppose that (1) and (2) hold. Note that (2) implies axiom Closure for W and axioms (VS1), (VS2), and (VS5)-(VS8) hold for W since $W \subseteq V$ and the operations are the same. By (1) we have that $\mathbf{0}_V \in W$. Since $\boldsymbol{x} + \mathbf{0}_V = \boldsymbol{x}$ for all $\boldsymbol{x} \in W$ (since $W \subseteq V$) axiom (VS3) holds for W. Let $\boldsymbol{x} \in W$. We know that $-\boldsymbol{x} \in V$ by axiom (VS4) applied to V. By corollary 3.2.3 we have that $-\boldsymbol{x} = (-1)\boldsymbol{x}$ and hence $-\boldsymbol{x} \in W$ by (2). Thus, axiom (VS4) holds for W which completes the proof that W is a vector space and hence a subspace of V. □

Let U, W be subspaces of a vector space V. We will prove that $U \cap W$ is a subspace of V. Since U and W are subspaces of V, we have that $\mathbf{0}_V \in U$ and $\mathbf{0}_V \in W$. Thus $\mathbf{0}_V \in U \cap W$. Next suppose $\boldsymbol{x}, \boldsymbol{y} \in U \cap W$ and $a, b \in \mathbb{R}$. Then $\boldsymbol{x}, \boldsymbol{y} \in U$ and $\boldsymbol{x}, \boldsymbol{y} \in W$. Since U and W are subspaces we have that $a\boldsymbol{x} + b\boldsymbol{y} \in U$ and $a\boldsymbol{x} + b\boldsymbol{y} \in W$. Thus $a\boldsymbol{x} + b\boldsymbol{y} \in U \cap W$ and by theorem 3.2 we have that $U \cap W$ is a subspace of V. We have justified the following lemma.

Lemma 3.3.2

Let U, W be subspaces of a vector space V. Then $U \cap W$ is a subspace of V.

Exercise 3.5

Prove that $W = \{\mathbf{0}_V\}$ is a subspace of any vector space V.

Exercise 3.6

Is $W = \left\{\boldsymbol{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 1\right\}$ a subspace of $\mathbb{R}^3$? Explain.

Exercise 3.7

Is $W = \{\boldsymbol{f} \in \mathcal{F}(\mathbb{R}, \mathbb{R}) : \boldsymbol{f}(0) = 0\}$ a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$? Explain.

Exercise 3.8

Let U, W be subspaces of a vector space V. Prove that if $U \subseteq W$, then U is a subspace of W.

Definition 3.3

Suppose U and W are subspaces of V. The *sum* of U and W, denoted $U + W$, is defined to be

$$U + W = \{\boldsymbol{u} + \boldsymbol{w} : \boldsymbol{u} \in U \text{ and } \boldsymbol{w} \in W\}$$

Exercise 3.9

Suppose U and W are subspaces of V.

(1) Prove that $U + W$ is a subspace of V.
(2) Prove that if T is any subspace of V that contains $U \cup W$, then T contains $U + W$. Thus, $U + W$ is the *smallest* subspace of V containing $U \cup W$.

Definition 3.4

A vector space V is called the *direct sum* of subspaces U and W if each vector in V can be *uniquely* written as $u + w$, where $u \in U$ and $w \in W$. We denote that V is the direct sum of U and W by writing $V = U \oplus W$.

Exercise 3.10

Suppose U and W are subspaces of V. Prove that $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$.

3.4. Null Space of a Matrix

Let $A \in \mathbb{R}^{m\times n}$. We consider the question of how many different solutions $\boldsymbol{x} \in \mathbb{R}^n$ a linear system $A\boldsymbol{x} = \boldsymbol{b}$ can have for a particular right-hand side $\boldsymbol{b} \in \mathbb{R}^m$. The key to answering this question is the corresponding homogenous equation $A\boldsymbol{x} = \mathbf{0}$.

For any matrix $A \in \mathbb{R}^{m\times n}$ we have that $A\mathbf{0} = \mathbf{0}$ but for some matrices the system $A\boldsymbol{x} = \mathbf{0}$ has a nonzero solution. For example, it is easy to check that

$$\begin{bmatrix}1 & 2 & 1\\ 2 & 2 & 0\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}1\\ -1\\ 1\end{bmatrix} = \begin{bmatrix}0\\ 0\\ 0\end{bmatrix} \tag{3.8}$$

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ such that $A\boldsymbol{x} = \boldsymbol{b}$ for some $\boldsymbol{x} \in \mathbb{R}^n$. Suppose there exists $\boldsymbol{y} \neq \mathbf{0}$ such that $A\boldsymbol{y} = \mathbf{0}$. For any scalar $c \in \mathbb{R}$ we have that

$$A(\boldsymbol{x} + c\boldsymbol{y}) = A\boldsymbol{x} + cA\boldsymbol{y} = \boldsymbol{b} + \mathbf{0} = \boldsymbol{b}$$

and thus $A\boldsymbol{x} = \boldsymbol{b}$ has an infinite number of solutions. For example, note that

$$\begin{bmatrix}1 & 2 & 1\\ 2 & 2 & 0\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}1\\ 1\\ 0\end{bmatrix} = \begin{bmatrix}3\\ 4\\ 2\end{bmatrix} \tag{3.9}$$

and that

$$\begin{bmatrix}1 & 2 & 1\\ 2 & 2 & 0\\ 1 & 1 & 0\end{bmatrix}\left(\begin{bmatrix}1\\ 1\\ 0\end{bmatrix} + \begin{bmatrix}1\\ -1\\ 1\end{bmatrix}\right) = \begin{bmatrix}1 & 2 & 1\\ 2 & 2 & 0\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}2\\ 0\\ 1\end{bmatrix} = \begin{bmatrix}3\\ 4\\ 2\end{bmatrix} \tag{3.10}$$

We summarize these ideas in the following lemma.

Lemma 3.4.1

Let $A \in \mathbb{R}^{m\times n}$. If there exists $\boldsymbol{x} \neq \mathbf{0}$ such that $A\boldsymbol{x} = \mathbf{0}$, then for all $\boldsymbol{b} \in \mathbb{R}^m$ the linear system $A\boldsymbol{x} = \boldsymbol{b}$ either has no solution or an infinite number of solutions–a unique solution is not possible.

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $A\boldsymbol{x} = \mathbf{0}$ has the unique solution $\boldsymbol{x} = \mathbf{0}$. Suppose that for some $\boldsymbol{b} \in \mathbb{R}^m$ we have that $A\boldsymbol{y} = \boldsymbol{b}$ and $A\boldsymbol{z} = \boldsymbol{b}$. Subtracting the equations gives

$$\mathbf{0} = A\boldsymbol{y} - A\boldsymbol{z} = A(\boldsymbol{y} - \boldsymbol{z})$$

But since $A\boldsymbol{x} = \mathbf{0}$ has the unique solution $\boldsymbol{x} = \mathbf{0}$, we have that $\boldsymbol{y} - \boldsymbol{z} = \mathbf{0}$ and hence $\boldsymbol{y} = \boldsymbol{z}$. We have justified the following lemma.

Lemma 3.4.2

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $A\boldsymbol{x} = \mathbf{0}$ has the unique solution $\boldsymbol{x} = \mathbf{0}$. Then for all $\boldsymbol{b} \in \mathbb{R}^m$ the linear system $A\boldsymbol{x} = \boldsymbol{b}$ either has no solution or a single unique solution.

Definition 3.5

Let $A \in \mathbb{R}^{m\times n}$. We define the *null space* of A, denoted $N(A)$, as

$$N(A) = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \mathbf{0}\} \tag{3.11}$$

Let $A \in \mathbb{R}^{m\times n}$. We make two observations.

(1) Since $A\mathbf{0} = \mathbf{0}$ we have that $\mathbf{0} \in N(A)$.
(2) Suppose $\boldsymbol{x}, \boldsymbol{y} \in N(A)$ and let $a, b \in \mathbb{R}$. Then we have $A\boldsymbol{x} = \mathbf{0}$ and $A\boldsymbol{y} = \mathbf{0}$. Hence

$$A(a\boldsymbol{x} + b\boldsymbol{y}) = aA\boldsymbol{x} + bA\boldsymbol{y} + = a\mathbf{0} + b\mathbf{0} = \mathbf{0}$$

and $a\boldsymbol{x} + b\boldsymbol{y}$ is in $N(A)$.

Thus by theorem 3.2 we have the following result.

Lemma 3.4.3

Let $A \in \mathbb{R}^{m\times n}$. Then $N(A)$ is a subspace of $\mathbb{R}^n$.

For any $a, b, c \in \mathbb{R}$, the plane

$$W = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : ax_1 + bx_2 + cx_3 = 0 \right\}$$

is a subspace of $\mathbb{R}^3$ since $W = N(A)$ where $A \in \mathbb{R}^{1\times 3}$ is given by $A = \begin{bmatrix} a & b & c \end{bmatrix}$.

Exercise 3.11

Is $W = \{\boldsymbol{x} \in \mathbb{R}^3 : x_1 = 3x_2 \text{ and } x_3 = -x_2\}$ a subspace of $\mathbb{R}^3$? Explain.

Note that

$$W_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x_1 - x_2 = 0 \right\}, \qquad W_2 = \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x_1 + x_2 = 0 \right\}$$

are subspaces of $\mathbb{R}^2$ since $W_1 = N(\begin{bmatrix} 1 & -1 \end{bmatrix})$ and $W_2 = N(\begin{bmatrix} 1 & 1 \end{bmatrix})$. Let $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Note that $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \in W_1$ and hence $W_1 \cup W_2$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix} \in W_2$ and hence $W_1 \cup W_2$. However, the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ is not in $W_1 \cup W_2$ and hence $W_1 \cup W_2$ is not a subspace of $\mathbb{R}^2$. We have shown that if U and W are subspaces of V, then $U \cup W$ is *not* always a subspace of V.

Exercise 3.12

Let U and W be subspaces of a vector space V. Prove that $U \cup W$ is a subspace of V if and only if $U \subseteq W$ or $W \subseteq U$.

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. By theorem 2.1, we have that $A\boldsymbol{x} = \boldsymbol{0}$ if and only if $C\boldsymbol{x} = \boldsymbol{0}$. We summarize this result in the following lemma.

Lemma 3.4.4

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. Then we have that

$$N(A) = N(C)$$

Consider the matrix

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

Adding -2 times the first row to the third row and fourth row gives the matrix

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

Multiplying the second row by $-1/2$ and then adding 2 times the second row to the third row and -2 times the second row to the fourth row gives the matrix

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

Since $C\boldsymbol{x} = \boldsymbol{0}$ if and only if $\boldsymbol{x} = \boldsymbol{0}$ we have that $N(C) = \{\boldsymbol{0}\}$ and hence $N(A) = \{\boldsymbol{0}\}$ by lemma 3.4.4. Thus for all $\boldsymbol{b} \in \mathbb{R}^m$, the linear system $A\boldsymbol{x} = \boldsymbol{b}$ either has no solution or a single unique solution by lemma 3.4.2.

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. If $m < n$, we call $A\boldsymbol{x} = \boldsymbol{b}$ an underdetermined linear system. It turns out that such linear systems can *never* have a single unique solution because there has to be a *nonzero* vector $\boldsymbol{x}$ in $N(A)$.

We first consider the case where $A \in \mathbb{R}^{m\times(m+1)}$. As a specific example, let $A \in \mathbb{R}^{3\times 4}$ be given by

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

and consider the problem of finding $\boldsymbol{x} \in N(A)$ such that $\boldsymbol{x} \neq \boldsymbol{0}$. We look for a smaller version of the problem within our problem. We first try to construct a solution to our problem using the fact that

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

We note that

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

This first attempt would have been successful if $a_{21} = 0$ and $a_{31} = 0$ since

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

This failed attempt shows us that we must first apply elementary row operations to zero out the second and third rows in the first column. Fortunately, we can do this without changing the set of solutions by theorem 2.1. Adding -2 times the first row to the second row and -1 times the first row to the third row of A gives

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

It is easy to check that

$$\begin{bmatrix} -2 & 1 & -3 \\ -1 & 2 & -3 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{3.12}$$

and that

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

Finally we verify that

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

as expected.

Notice that to make this process work we had to be *given* a solution to the smaller problem (3.12). On the other hand that smaller problem contains a still smaller problem and so on. This problem reduction idea can be turned into a formal proof using mathematical induction on m. The induction begins with $m = 1$. Suppose $A = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$. If $a_{11} = 0$ set $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, otherwise set $\boldsymbol{x} = \begin{bmatrix} -a_{12}/a_{11} \\ 1 \end{bmatrix}$. In either case $\boldsymbol{x}$ is a nonzero solution to $A\boldsymbol{x} = \mathbf{0}$. Now suppose the result is true for some integer $m \geq 1$. Let $A \in \mathbb{R}^{(m+1)\times(m+2)}$ be arbitrary. We will prove that $A\boldsymbol{x} = \mathbf{0}$ has a nonzero solution. If the first column of A is $\mathbf{0}$, then $A\boldsymbol{e}_1$ and the result holds. Otherwise we can perform elementary row operations on A to find a matrix B of the form

$$B = \begin{bmatrix} 1 & \boldsymbol{c}^T \\ \mathbf{0} & D \end{bmatrix}$$

where $\boldsymbol{c} \in \mathbb{R}^{m+1}$ and $D \in \mathbb{R}^{m\times(m+1)}$. By the inductive hypothesis, there exists $\boldsymbol{y} \in \mathbb{R}^{m+1}$ such that $\boldsymbol{y} \neq \mathbf{0}$ and $D\boldsymbol{y} = \mathbf{0}$. Define

$$\boldsymbol{x} = \begin{bmatrix} -c_1y_1 - \cdots - c_{m+1}y_{m+1} \\ y_1 \\ \vdots \\ y_{m+1} \end{bmatrix}$$

and note that $\boldsymbol{x} \in \mathbb{R}^{m+2}$, $\boldsymbol{x} \neq 0$ and $B\boldsymbol{x} = \mathbf{0}$. Since B was obtained by performing elementary row operations on A we also have $A\boldsymbol{x} = \mathbf{0}$ by theorem 2.1. We have justified the following lemma.

Lemma 3.4.5

For all $m \geq 1$ and $A \in \mathbb{R}^{m\times(m+1)}$ there exists $\boldsymbol{x} \in N(A)$ such that $\boldsymbol{x} \neq \mathbf{0}$.

Next let's consider the general underdetermined homogenous system $A\boldsymbol{x} = \mathbf{0}$ where $A \in \mathbb{R}^{m\times n}$ where $m < n$. If $n = m + 1$, then $A\boldsymbol{x} = \mathbf{0}$ for some $\boldsymbol{x} \neq \mathbf{0}$ by lemma 3.4.5. If not, we have $n > m + 1$ and can write A in the form

$$A = \begin{bmatrix} B & C \end{bmatrix}$$

where $B \in \mathbb{R}^{m\times(m+1)}$ and $C \in \mathbb{R}^{m\times(n-m-1)}$. By lemma 3.4.5, there exists $\boldsymbol{y} \in \mathbb{R}^{m+1}$ such that $\boldsymbol{y} \neq \mathbf{0}$ and $B\boldsymbol{y} = \mathbf{0}$. Define $\boldsymbol{x} \in \mathbb{R}^n$ by

$$\boldsymbol{x} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and note that $A\boldsymbol{x} = \mathbf{0}$ where $\boldsymbol{x} \neq \mathbf{0}$. We have justified the next theorem.

Theorem 3.3

Let $A \in \mathbb{R}^{m\times n}$. If $m < n$, then there exists $\boldsymbol{x} \in N(A)$ such that $\boldsymbol{x} \neq \mathbf{0}$.

The following corollary follows from theorem 3.3 and lemma 3.4.1

Corollary 3.4.1

Let $A \in \mathbb{R}^{m\times n}$. If $m < n$, then for all $\boldsymbol{b} \in \mathbb{R}^m$ the linear system $A\boldsymbol{x} = \boldsymbol{b}$ either has no solution or an infinite number of solutions–a unique solution is not possible.

CHAPTER 4

Span and Linear Independence

4.1. Linear Combinations and Span

Definition 4.1

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be vectors in a vector space V. We say that a vector $\boldsymbol{v} \in V$ is a *linear combination* of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ if there exist scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$ such that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

Consider the vectors $\boldsymbol{v}, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in \mathbb{R}^3$ given by

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ -4 \\ -2 \end{bmatrix}, \qquad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Since $\boldsymbol{v} = 2\boldsymbol{v}_1 - 3\boldsymbol{v}_2 + 2\boldsymbol{v_3}$, the vector $\boldsymbol{v}$ is a linear combination of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

Exercise 4.1

Suppose $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$. Prove that for all $\boldsymbol{b} \in \mathbb{R}^m$ the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if $\boldsymbol{b}$ is a linear combination of the column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

Consider the vectors $\boldsymbol{v}, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ in $\mathbb{R}^3$ given by

$$\boldsymbol{v} = \begin{bmatrix} 5 \\ 8 \\ 6 \end{bmatrix}, \quad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ -1 \\ -3 \end{bmatrix}, \quad \boldsymbol{v}_4 = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix}$$

Exercise 4.1 gives that $\boldsymbol{v}$ is a linear combination of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ if and only if

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

for some $x \in \mathbb{R}^4$ where

$$A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 3 & -1 & 4 \\ 1 & 4 & -3 & -3 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 5 \\ 8 \\ 6 \end{bmatrix}$$

We consider the augmented matrix

$$A|\boldsymbol{b} = \left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 5 \\ 2 & 3 & -1 & 4 & 8 \\ 1 & 4 & -3 & -3 & 6 \end{array}\right]$$

Row reducing gives

$$C|\boldsymbol{d} = \left[\begin{array}{cccc|c} 1 & 0 & 1 & 5 & 0 \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Thus, the system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent and hence $\boldsymbol{v}$ is *not* a a linear combination of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$.

Exercise 4.2

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be vectors in a vector space V. Prove that the *set* of linear combinations of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ is a *subspace* of V.

Definition 4.2

Let S be a nonempty set of vectors in a vector space V. A vector $\boldsymbol{v}$ is called a *linear combination* of vectors in S if there exist vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in S$ and scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$ such that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

Let $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 6 \\ -3 \end{bmatrix} \right\}$ be a set of vectors in $\mathbb{R}^3$. Then

$$\boldsymbol{v} = 3\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} - 2\begin{bmatrix} -2 \\ 6 \\ -3 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + 7\begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}$$

is a linear combination of vectors in S.

Definition 4.3

The *span* of a nonempty set S in a vector space V, denoted $\operatorname{span}(S)$, is the set of all linear combinations of vectors in S. We define $\operatorname{span}(\varnothing) = \{0\}$.

Let S be a set of vectors in a vector space V. For all $\boldsymbol{v} \in S$ we have $\boldsymbol{v} = 1\boldsymbol{v}$ and thus

$$\boxed{S \subseteq \operatorname{span}(S)} \tag{4.1}$$

Let V be a vector space. Note that $\operatorname{span}(\varnothing) = \{\boldsymbol{0}\}$ is a subspace of V. Next, let S be a nonempty set in V. We make two observations.

(1) Let $\boldsymbol{v} \in S$. Then $\boldsymbol{0} = 0\boldsymbol{v}$ and we have that $\boldsymbol{0} \in \operatorname{span}(S)$.
(2) Let $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{span}(S)$ and $a, b \in \mathbb{R}$. Then we have

$$\boldsymbol{x} = c_1\boldsymbol{w}_1 + \cdots + c_m\boldsymbol{w}_m, \qquad \boldsymbol{v}_2 = d_1\boldsymbol{u}_1 + \cdots + d_n\boldsymbol{u}_n$$

for some vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \in S$ and $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n} \in S$ as well as scalars $c_1, \ldots, c_m \in \mathbb{R}$ and $d_1, \ldots, d_n \in \mathbb{R}$. Thus we have that

$$a\boldsymbol{x} + b\boldsymbol{y} = (ac_1)\boldsymbol{w}_1 + \cdots + (ac_m)\boldsymbol{w}_m + (bd_1)\boldsymbol{u}_1 + \cdots + (bd_n)\boldsymbol{u}_n$$

is a linear combination of vectors in S. Hence $a\boldsymbol{x} + b\boldsymbol{y}$ is in $\operatorname{span}(S)$.

Thus, by theorem 3.2 we have that $\operatorname{span}(S)$ is a subspace of V. Next let S be a subset of a vector space V and let W be a subspace of V such that $S \subseteq W$. If $S = \varnothing$, then $\operatorname{span}(S) = \operatorname{span}(\varnothing) = \{\boldsymbol{0}\}$ is a subset of W since W is a subspace. Next, suppose S is nonempty and let $\boldsymbol{v} \in \operatorname{span}(S)$. Then

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are vectors in S and also in W since $S \subseteq W$. Since W is a subspace, $\boldsymbol{v} \in W$ since W is closed under addition and scalar multiplication. We summarize our results in the next theorem.

Theorem 4.1

Let S be a subset of a vector space V.

(1) $S \subseteq \operatorname{span}(S)$
(2) $\operatorname{span}(S)$ is a subspace of V.
(3) If W is a subspace of V and $S \subseteq W$, then $\operatorname{span}(S) \subseteq W$.

Suppose S and T are subsets of a vector space V. If $S \subseteq \text{span}(T)$, then

$$\text{span}(S) \subseteq \text{span}(T)$$

by theorem 4.1. We have justified the following.

Corollary 4.1.1

Suppose S and T are subsets of a vector space V. If $S \subseteq \text{span}(T)$, then

$$\text{span}(S) \subseteq \text{span}(T)$$

Consider the subsets of $\mathbb{R}^2$ given by

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \qquad T = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

Since

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

we have that $S \subseteq \text{span}(T)$ and hence

$$\text{span}(S) \subseteq \text{span}(T) \tag{4.2}$$

by corollary 4.1.1. Since

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

we have that $T \subseteq \text{span}(S)$ and hence

$$\text{span}(T) \subseteq \text{span}(S) \tag{4.3}$$

by corollary 4.1.1. Combing equations (4.2) and (4.3) gives

$$\text{span}(S) = \text{span}(T)$$

Exercise 4.3

Let $\boldsymbol{u}, \boldsymbol{v}$ be vectors in a vector space V. Prove that

$$\text{span}(\{\boldsymbol{u}, \boldsymbol{v}\}) = \text{span}(\{\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v}\})$$

Exercise 4.4

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be vectors in a vector space V and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be *nonzero* scalars in $\mathbb{R}$. Prove that

$$\text{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}) = \text{span}(\{\alpha_1\boldsymbol{v}_1, \alpha_2\boldsymbol{v}_2, \ldots, \alpha_n\boldsymbol{v}_n\})$$

Exercise 4.5

Suppose S and T are subsets of a vector space V. Prove that if $S \subseteq T$, then

$$\text{span}(S) \subseteq \text{span}(T)$$

Exercise 4.6

Consider the set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of vectors in a vector space V. Prove that $\text{span}(S)$ is equal to the set of linear combinations of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$.

Exercise 4.7

Consider the vectors $\boldsymbol{v}, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ in $\mathbb{R}^3$ given by

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ 2 \\ 11 \end{bmatrix}, \quad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ -1 \\ -3 \end{bmatrix}, \quad \boldsymbol{v}_4 = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix}$$

Is $\boldsymbol{v} \in \operatorname{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\})$? Explain.

Definition 4.4

Let S be a subset of a vector space V. We say that S *generates* V if

$$\operatorname{span}(S) = V$$

We also call S a *generating set* for V.

The set of column vectors of I_n given by $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ *generates* $\mathbb{R}^n$. Let S be a subset of a vector space V. Then S is a *generating set* for $\operatorname{span}(S)$.

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 & -4 \\ -1 & 1 & -4 & 2 \\ 1 & 2 & 1 & -5 \end{bmatrix}$$

We will find a generating set for $N(A)$, the null space of A. We apply Guass-Jordan elimination to reduce A to the matrix C in reduced row echelon form

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

Since $N(A) = N(C)$ by lemma 3.4.4 it suffices to find a generating set for $N(C)$. We rewrite the system $C\boldsymbol{x} = \mathbf{0}$ as

$$x_1 + 3x_3 - 3x_4 = 0, \qquad x_2 - x_3 - x_4 = 0$$

After solving for left most variable in each equation we get

$$x_1 = -3x_3 + 3x_4, \qquad x_2 = x_3 + x_4$$

We next let $t_1 = x_3$ and $t_2 = x_4$ and note that $C\boldsymbol{x} = \mathbf{0}$ if and only if

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3t_1 + 3t_2 \\ t_1 + t_2 \\ t_1 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Thus $C\boldsymbol{x} = \mathbf{0}$ if and only if $\boldsymbol{x}$ is a *linear combination* of the vectors

$$\boldsymbol{v_1} = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v_2} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Hence we have

$$N(A) = N(C) = \operatorname{span}(\{\boldsymbol{v_1}, \boldsymbol{v_2}\})$$

and we say that $N(A)$ is *generated* by the set $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$.

Exercise 4.8

Consider the matrix

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

Find a generating set for $N(A)$.

4.2. Column Space of a Matrix

Definition 4.5

Suppose $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. We define the column space of A, denoted $\operatorname{col}(A)$, as

$$\operatorname{col}(A) = \operatorname{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\})$$

Exercise 4.9

Let $A \in \mathbb{R}^{m\times n}$.

(1) Prove that

$$\operatorname{col}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} = \{\boldsymbol{b} \in \mathbb{R}^m : A\boldsymbol{x} = \boldsymbol{b} \text{ for some } \boldsymbol{x} \in \mathbb{R}^n\}$$

(2) Prove that the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for *all* $\boldsymbol{b} \in \mathbb{R}^m$ if and only if $\operatorname{col}(A) = \mathbb{R}^m$.

The next corollary follows immediately from theorem 4.1.

Corollary 4.2.1

Let $A \in \mathbb{R}^{m\times n}$. Then $\operatorname{col}(A)$ is a subspace of $\mathbb{R}^m$.

If $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, then the set of column vectors $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ *generates* $\operatorname{col}(A)$.

Exercise 4.10

Let $A \in \mathbb{R}^{4\times 4}$. Suppose you are given that

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

Find $\boldsymbol{x}$ such that $A\boldsymbol{x} = \begin{bmatrix} 2 \\ 0 \\ 4 \\ 6 \end{bmatrix}$ and find $\boldsymbol{y}$ such that $A\boldsymbol{y} = \begin{bmatrix} 1 \\ -1 \\ 5 \\ 4 \end{bmatrix}$.

Consider the matrices

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

with column vectors $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\}$ and $T = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \boldsymbol{b}_4, \boldsymbol{b}_5\}$ respectively. Since the sets S and T contain the same vectors, they are equal and thus

$$\operatorname{col}(A) = \operatorname{span}(S) = \operatorname{span}(T) = \operatorname{col}(B)$$

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. In general, we will *not* have that $\text{col}(A) = \text{col}(C)$. For example, let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Add -1 times the first row to the second row to get $C = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Note that the vector $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is in $\text{col}(A)$ since $A\boldsymbol{e}_1 = \boldsymbol{b}$. However, $\boldsymbol{b}$ is not in $\text{col}(C)$ since $C\boldsymbol{x} = \boldsymbol{b}$ has no solution. The next theorem covers a *special case* where we do have $\text{col}(A) = \text{col}(C)$.

Theorem 4.2

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. Then $\text{col}(A) = \mathbb{R}^m$ if and only if $\text{col}(C) = \mathbb{R}^m$.

PROOF. Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. By theorem 1.1 there exists an invertible matrix $P \in \mathbb{R}^{m\times m}$ such that $C = PA$. First suppose that $\text{col}(A) = \mathbb{R}^m$ and let $\boldsymbol{b} \in \mathbb{R}^m$. Then $A\boldsymbol{x} = P^{-1}\boldsymbol{b}$ for some $\boldsymbol{x} \in \mathbb{R}^n$ and hence we have that

$$\boldsymbol{b} = PA\boldsymbol{x} = C\boldsymbol{x}$$

which implies that $\boldsymbol{b} \in \text{col}(C)$ and thus $\text{col}(C) = \mathbb{R}^m$. Next suppose that $\text{col}(C) = \mathbb{R}^m$ and let $\boldsymbol{b} \in \mathbb{R}^m$. Then $C\boldsymbol{x} = P\boldsymbol{b}$ for some $\boldsymbol{x} \in \mathbb{R}^n$ and hence we have that

$$\boldsymbol{b} = P^{-1}C\boldsymbol{x} = A\boldsymbol{x}$$

which implies that $\boldsymbol{b} \in \text{col}(A)$ and thus $\text{col}(A) = \mathbb{R}^m$. □

Exercise 4.11

Consider the matrix

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

Does $Ax = b$ have a solution for all $b \in \mathbb{R}^3$? Explain.

Exercise 4.12

Let $A \in \mathbb{R}^{m\times n}$. Let $C \in \mathbb{R}^{m\times n}$ be a matrix resulting from applying any number of elementary row operations to A. Prove that

$$\text{col}(A^T) = \text{col}(C^T)$$

4.3. Linearly Independent Sets

Consider the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ in $\mathbb{R}^3$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

Since $\boldsymbol{v}_3 = \boldsymbol{v}_1 + \boldsymbol{v}_2$ any linear combination of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3$$

is *also* a linear combination of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ since

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3(\boldsymbol{v}_1 + \boldsymbol{v}_2) = (c_1 + c_3)\boldsymbol{v}_1 + (c_2 + c_3)\boldsymbol{v}_2$$

Let S be a set of vectors in a vector space V and suppose that for some $\boldsymbol{v} \in S$ we have that

$$\boldsymbol{v} \in \text{span}(S - \{\boldsymbol{v}\})$$

First, since $S - \{\boldsymbol{v}\} \subseteq S$ we have that

$$\text{span}(S - \{\boldsymbol{v}\}) \subseteq \text{span}(S)$$

by exercise 4.5.

Next, since $S - \{\boldsymbol{v}\} \subseteq \text{span}(S - \{\boldsymbol{v}\})$ by theorem 4.1 and and $\boldsymbol{v} \in \text{span}(S - \{\boldsymbol{v}\})$ we have that

$$S \subseteq \text{span}(S - \{\boldsymbol{v}\})$$

and hence

$$\text{span}(S) \subseteq \text{span}(S - \{\boldsymbol{v}\})$$

by corollary 4.1.1. We have justified the following result.

Lemma 4.3.1

Let S be a set of vectors in a vector space V and suppose that for some $\boldsymbol{v} \in S$ we have

$$\boldsymbol{v} \in \text{span}(S - \{\boldsymbol{v}\})$$

Then we have that

$$\text{span}(S) = \text{span}(S - \{\boldsymbol{v}\})$$

Let $A \in \mathbb{R}^{3\times 4}$ be given by

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

and consider the column space of A

$$\text{col}(A) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\})$$

where

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{a}_2 = \begin{bmatrix} -1 \\ 1 \\ -4 \end{bmatrix}, \qquad \boldsymbol{a}_3 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \qquad \boldsymbol{a}_4 = \begin{bmatrix} 0 \\ 4 \\ -4 \end{bmatrix}$$

The set $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\}$ *generates* $\text{col}(A)$ but it is possible to generate $\text{col}(A)$ with *fewer* than 4 vectors. To see how, we apply Guass-Jordan elimination to reduce A to the matrix C in reduced row echelon form

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

with column vectors

$$\boldsymbol{c}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{c}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{c}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{c}_4 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$$

By inspection we observe that

$$2\boldsymbol{c}_1 + \boldsymbol{c}_2 - \boldsymbol{c}_3 = \boldsymbol{0}$$

But then we must also have that

$$2\boldsymbol{a}_1 + \boldsymbol{a}_2 - \boldsymbol{a}_3 = \boldsymbol{0}$$

or

$$\boldsymbol{a}_3 = 2\boldsymbol{a}_1 + \boldsymbol{a}_2 \tag{4.4}$$

since elementary row operations preserve solutions to $A\boldsymbol{x} = \boldsymbol{0}$ by theorem 2.1. By lemma 4.3.1 we have that

$$\text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\}) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_4\})$$

Similarily, we note that

$$2\boldsymbol{c}_1 + 2\boldsymbol{c}_2 - \boldsymbol{c}_4 = \boldsymbol{0}$$

and hence

$$\boldsymbol{a}_4 = 2\boldsymbol{a}_1 + 2\boldsymbol{a}_2 \tag{4.5}$$

By lemma 4.3.1 we have

$$\text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_4\}) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2\})$$

and we have shown that

$$\text{col}(A) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2\})$$

Similarily, It is easy to verify that

$$\boldsymbol{a}_1 = \boldsymbol{a}_3 - \frac{\boldsymbol{a}_4}{2}, \qquad \boldsymbol{a}_2 = -\boldsymbol{a}_3 + \boldsymbol{a}_4$$

and hence

$$\text{col}(A) = \text{span}(\{\boldsymbol{a}_3, \boldsymbol{a}_4\})$$

Definition 4.6

Let S be a set of vectors in a vector space V. We say that S is *linearly dependent* if there exists a vector $\boldsymbol{v} \in S$ such that $\boldsymbol{v} \in \text{span}(S - \{\boldsymbol{v}\})$. We say that S is *linearly independent* if it is *not* linearly dependent.

The empty set $\varnothing$ is linearly independent by definition since it contains no vectors. The set $S = \{\mathbf{0}\}$ is linearly dependent since

$$\text{span}(\{\mathbf{0}\}) = \{\mathbf{0}\} = \text{span}(\varnothing)$$

The set $S = \{\boldsymbol{v}\}$ where $\boldsymbol{v} \neq \mathbf{0}$ is linearly independent since in this case $\text{span}(\{\boldsymbol{v}\}) \neq \{\mathbf{0}\} = \text{span}(\varnothing)$.

Consider the set of vectors $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ in $\mathbb{R}^3$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

Since $\boldsymbol{v}_1 = \boldsymbol{v}_2 - \boldsymbol{v}_3$, we have that $\boldsymbol{v}_1 \in \text{span}(S - \{\boldsymbol{v}_1\})$ and S is linearly dependent.

Exercise 4.13

Let S and T be subsets of a vector space V and suppose that $S \subseteq T$.

(1) Prove that if S is linearly dependent, then T is linearly dependent.
(2) Prove that if T is linearly independent, then S is linearly independent.

The next theorem gives our primary test for linear independence. We leave the proof as an exercise.

Theorem 4.3

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be vectors in a vector space V. Then $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_n\}$ is a linearly independent set of n vectors if and only if

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = \mathbf{0}$$

has the unique solution $c_1 = c_2 = \cdots = c_n = 0$.

Exercise 4.14

Prove thereom 4.3.

Consider the set $S = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ of column vectors of I_n. If

$$c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + \cdots + c_n\boldsymbol{e}_n = \mathbf{0}$$

then

$$\boldsymbol{c} = I_n\boldsymbol{c} = c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + \cdots + c_n\boldsymbol{e}_n = \mathbf{0} = \mathbf{0}$$

Thus S is a a linearly independent set of n vectors in $\mathbb{R}^n$ by theorem 4.3.

Again consider the matrix

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

with column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4$. We have shown that

$$\text{col}(A) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2\})$$

Suppose that

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 = \boldsymbol{0}$$

which we can write in matrix form as

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & -4 \end{bmatrix} \boldsymbol{c} = \boldsymbol{0}$$

Applying Gauss-Jordan elimination gives us the system

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{c} = \boldsymbol{0}$$

which has the unique solution $\boldsymbol{c} = \boldsymbol{0}$. Thus, the set $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is linearly independent by theorem 4.3. It is easy to prove that the set $T = \{\boldsymbol{a}_3, \boldsymbol{a}_4\}$ is also linearly independent. These examples suggest that any generating set for $\text{col}(A)$ must contain *at least* two vectors.

Exercise 4.15

Suppose $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Prove that $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ is a linearly independent set of n vectors if and only if $N(A) = \{\boldsymbol{0}\}$.

Exercise 4.16

Consider the following vectors in $\mathbb{R}^3$.

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Is the set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ linearly independent? Explain.

Exercise 4.17

Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a linearly independent set of n vectors in $\mathbb{R}^m$. Prove that $n \leq m$.

Let $S = \{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$ be a linearly independent set of 3 vectors in $\mathbb{R}^k$ and consider the set

$$T = \{\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{w}, \boldsymbol{v} + \boldsymbol{w}\}$$

Suppose

$$c_1(\boldsymbol{u} + \boldsymbol{v}) + c_2(\boldsymbol{u} + \boldsymbol{w}) + c_3(\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{0}$$

After collecting terms we get

$$(c_1 + c_2)\boldsymbol{u} + (c_1 + c_3)\boldsymbol{v} + (c_2 + c_3)\boldsymbol{w} = \boldsymbol{0}$$

Using theorem 4.3 we note that

$$c_1 + c_2 = 0, \qquad c_1 + c_3 = 0, \qquad c_2 + c_3 = 0$$

which in matrix form says that $\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}\boldsymbol{c} = \boldsymbol{0}$. Applying Gauss-Jordan elimination gives us the system $\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\boldsymbol{c} = \boldsymbol{0}$ which has the unique solution $\boldsymbol{c} = \boldsymbol{0}$. Thus T is a linearly independent set of 3 vectors by theorem 4.3.

Exercise 4.18

Suppose $S = \{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$ is a linearly independent set of 3 vectors in a vector space V. Prove that

$$T = \{\boldsymbol{u} + \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{u} - \boldsymbol{v} + \boldsymbol{w}, -\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}\}$$

is also a linearly independent set of 3 vectors in V.

Exercise 4.19

Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$. Suppose that

$$\boldsymbol{f}(t) = \sin(t), \qquad \boldsymbol{g}(t) = \cos(t)$$

Is the set of vectors $S = \{\boldsymbol{f}, \boldsymbol{g}\}$ linearly independent? Explain.

Exercise 4.20

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $N(A) = \{\boldsymbol{0}\}$. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a linearly independent set of k vectors in $\mathbb{R}^n$. Prove that the set $T = \{A\boldsymbol{v}_1, A\boldsymbol{v}_2, \ldots, A\boldsymbol{v}_k\}$ is a linearly independent set of k vectors in $\mathbb{R}^m$.

The next result will be a fundamental tool for obtaining an upper bound on the size of a linearly independent set in a vector space V and a lower bound on the size of a generating set for V.

Theorem 4.4

If $T = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n\}$ is a linearly independent subset of a vector space V containing n vectors and $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is a generating set for V containing m vectors, then $n \leq m$.

Proof. Since S generates V, for all i in $1, \ldots, n$ we have that

$$\boldsymbol{w}_i = a_{1i}\,\boldsymbol{v}_1 + \cdots + a_{mi}\,\boldsymbol{v}_m$$

for some scalars $a_{1i}, \ldots, a_{mi} \in \mathbb{R}$. Let $C \in \mathbb{R}^{m\times n}$ be the matrix

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

Let $\boldsymbol{x} \in \mathbb{R}^n$ be any solution to $C\boldsymbol{x} = \boldsymbol{0}$. We note that

$$\begin{aligned} x_1\boldsymbol{w}_1 + \cdots + x_n\boldsymbol{w}_n &= x_1(a_{11}\,\boldsymbol{v}_1 + \cdots + a_{m1}\boldsymbol{v}_m) + \cdots + x_n(a_{1n}\,\boldsymbol{v}_1 + \cdots + a_{mn}\boldsymbol{v}_m)\\ &= (x_1a_{11} + \cdots + x_na_{1n})\boldsymbol{v}_1 + \cdots + (x_1a_{m1} + \cdots + x_na_{mn})\boldsymbol{v}_m\\ &= 0\boldsymbol{v}_1 + \cdots + 0\boldsymbol{v}_m\\ &= 0 \end{aligned}$$

Since T is a linearly independent set of n vectors we have that $x_1 = x_2 = \cdots = x_n = 0$ by theorem 4.3 and thus $C\boldsymbol{x} = \boldsymbol{0}$ only if $\boldsymbol{x} = \boldsymbol{0}$. Finally, we note that $n \leq m$ by theorem 3.3. □

Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ be a generating set for $\mathbb{R}^n$ containing m vectors. Since $T = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is a linearly independent set of n vectors in $\mathbb{R}^n$ we must have $m \geq n$ by theorem 4.4.

Exercise 4.21

Let $A \in \mathbb{R}^{m\times n}$. Prove that if $m > n$, then there exists a vector $\boldsymbol{b} \in \mathbb{R}^m$ such that $A\boldsymbol{x} = \boldsymbol{b}$ does not have a solution.

Exercise 4.22

Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a subset of $\mathbb{R}^m$ containing n vectors. Prove that if S generates $\mathbb{R}^m$ and S is linearly independent, then $n = m$.

Exercise 4.23

Let $A \in \mathbb{R}^{m\times n}$. Prove that if $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for all $\boldsymbol{b} \in \mathbb{R}^m$, then $m = n$.

4.4. Invertibility and the Vandermonde Matrix

Let $A \in \mathbb{R}^{n\times n}$ and suppose that the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for all $\boldsymbol{b} \in \mathbb{R}^n$. Then $A\boldsymbol{x} = \boldsymbol{0}$ must have the unique solution $\boldsymbol{x} = \boldsymbol{0}$. It turns out that $A\boldsymbol{x} = \boldsymbol{0}$ having the unique solution $\boldsymbol{x} = \boldsymbol{0}$ is not only a necessary condition to guarantee a unique solution for all $\boldsymbol{b} \in \mathbb{R}^n$, but it is also a sufficient condition. Let $A \in \mathbb{R}^{n\times n}$ and suppose that the linear system $A\boldsymbol{x} = \boldsymbol{0}$ has the unique solution $\boldsymbol{x} = \boldsymbol{0}$. By lemma 3.4.2, it remains to show that $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for all $\boldsymbol{b} \in \mathbb{R}^n$. By theorem 3.3, the linear system $(A|b)\boldsymbol{y} = \boldsymbol{0}$ has a nonzero solution $\boldsymbol{y} \neq \boldsymbol{0}$. Thus we have

$$y_1\boldsymbol{a}_1 + y_2\boldsymbol{a}_2 + \cdots + y_n\boldsymbol{a}_n + y_{n+1}\boldsymbol{b} = 0 \tag{4.6}$$

where $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ are the column vectors of A and at least one of $y_1, y_2, \ldots, y_{n+1}$ is nonzero. But since $A\boldsymbol{x} = \boldsymbol{0}$ only if $\boldsymbol{x} = \boldsymbol{0}$ we must have $y_{n+1} \neq 0$. Solving for $\boldsymbol{b}$ in equation (4.6) gives

$$\frac{-y_1}{y_{n+1}}\boldsymbol{a}_1 + \frac{-y_2}{y_{n+1}}\boldsymbol{a}_2 + \cdots + \frac{-y_n}{y_{n+1}}\boldsymbol{a}_n = \boldsymbol{b} \tag{4.7}$$

and thus

$$\boldsymbol{x} = \begin{bmatrix} \frac{-y_1}{y_{n+1}} \\ \frac{-y_2}{y_{n+1}} \\ \vdots \\ \frac{-y_n}{y_{n+1}} \end{bmatrix}$$

is a solution to $A\boldsymbol{x} = \boldsymbol{b}$. We have justified the following result.

Lemma 4.4.1

Let $A \in \mathbb{R}^{n\times n}$. Then $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for all $\boldsymbol{b} \in \mathbb{R}^n$ if and only if $N(A) = \{\boldsymbol{0}\}$.

Exercise 4.24

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 0 & 5 & -4 \end{bmatrix}$. Does $A\boldsymbol{x} = \boldsymbol{b}$ have a unique solution for all $\boldsymbol{b} \in \mathbb{R}^3$?

Let $A, B \in \mathbb{R}^{n\times n}$. In general we cannot say that $AB = BA$ since matrix multiplication is not commutative. It turns out, however, that if $AB = I_n$ then we must have $BA = I_n$ as well. To show this suppose that $AB = I_n$. Let $\boldsymbol{x} \in \mathbb{R}$ be any solution to $B\boldsymbol{x} = \boldsymbol{0}$ and note that

$$\boldsymbol{x} = I_n\boldsymbol{x} = (AB)\boldsymbol{x} = A(B\boldsymbol{x}) = A\boldsymbol{0} = \boldsymbol{0}$$

Thus, $B\boldsymbol{x} = \boldsymbol{0}$ only if $\boldsymbol{x} = \boldsymbol{0}$ and by lemma 4.4.1 we have that $B\boldsymbol{x} = \boldsymbol{y}$ has a unique solution for all $\boldsymbol{y} \in \mathbb{R}^n$. Next, for an arbitrary $i \in 1, \ldots, n$ choose $\boldsymbol{x} \in \mathbb{R}^n$ such that $B\boldsymbol{x} = \boldsymbol{e}_i$ and note that

$$(BA)\boldsymbol{e}_i = (BA)(B\boldsymbol{x}) = B((AB)\boldsymbol{x}) = B(I_n\boldsymbol{x}) = B\boldsymbol{x} = \boldsymbol{e}_i$$

Since $i \in 1, \ldots, n$ was arbitrary we have that $BA = I_n$ by exercise 1.2. We have justified the following lemma.

Lemma 4.4.2

Let $A, B \in \mathbb{R}^{n\times n}$. If $AB = I_n$, then $BA = I_n$.

Let $A \in \mathbb{R}^{n\times n}$ and suppose that $A\boldsymbol{x} = \boldsymbol{e}_i$ has a solution for all $i \in 1, \ldots, n$. For i in $1, \ldots, n$ choose $\boldsymbol{b}_i \in \mathbb{R}^n$ such that $A\boldsymbol{b}_i = \boldsymbol{e}_i$. Let $B \in \mathbb{R}^{n\times n}$ have column vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ and note that

$$AB = \begin{bmatrix} A\boldsymbol{b}_1 & \cdots & A\boldsymbol{b}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n \end{bmatrix} = I_n$$

Finally, we note that $BA = I_n$ by lemma 4.4.2 and hence A is invertible. We summarize our results with the following lemma.

Lemma 4.4.3

Let $A \in \mathbb{R}^{n\times n}$. If $A\boldsymbol{x} = \boldsymbol{e}_i$ has a solution for all i in $1, \ldots, n$, then A is invertible.

The following theorem gives several statements equivalent to invertibility of a matrix $A \in \mathbb{R}^{n\times n}$. We leave the proof as an exercise.

Theorem 4.5

Let $A \in \mathbb{R}^{n\times n}$. Then the following statements are equivalent.

(1) A is invertible.
(2) $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for all $\boldsymbol{b} \in \mathbb{R}^n$.
(3) $N(A) = \{\boldsymbol{0}\}$.
(4) $\mathrm{col}(A) = \mathbb{R}^n$.
(5) $A\boldsymbol{x} = \boldsymbol{e}_i$ has a solution for all i in $1, \ldots n$.

Exercise 4.25

Prove theorem 4.5.

We next return to the polynomial interpolation problem. Let $c_1, c_2, \ldots, c_n$ be *distinct* scalars in $\mathbb{R}$ and $b_1, b_2, \ldots, b_n$ be any scalars in $\mathbb{R}$. The polynomial interpolation problem is that of finding a polynomial of the form

$$f(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$$

such that $f(c_i) = b_i$ for all i in $1, \ldots, n$. The given constraints yield the linear system

$$\begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{4.8}$$

Define the matrix $A \in \mathbb{R}^{n\times n}$, called a Vandermonde matrix, as

$$A = \begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix}$$

We will show that A is invertible and hence the linear system (4.8) (and the associated polynomial interpolation problem) always has a unique solution. By theorem 4.5 it suffices to show that $A\boldsymbol{x} = \boldsymbol{e}_i$ has a solution for all i in $1, \ldots, n$. To this end we define the Lagrange polynomials $\boldsymbol{f_1}, \boldsymbol{f_2}, \ldots, \boldsymbol{f_n}$ by

$$f_i(t) = \frac{(t-c_1)\cdots(t-c_{i-1})(t-c_{i+1})\cdots(t-c_n)}{(c_i-c_1)\cdots(c_i-c_{i-1})(c_i-c_{i+1})\cdots(c_i-c_n)}$$

Note that $f_i(c_i) = 1$, and $f_i(c_k) = 0$ for $i \neq k$. Finally we write

$$f_i(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$$

for some scalars $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and note that

$$A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} f_i(c_1) \\ f_i(c_2) \\ \vdots \\ f_i(c_n) \end{bmatrix} = \boldsymbol{e}_i$$

We have justified the following result.

Corollary 4.4.1

Let $c_1, c_2, \ldots, c_n$ be *distinct* scalars in $\mathbb{R}$. The Vandermonde matrix $A \in \mathbb{R}^{n\times n}$ given by

$$A = \begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix}$$

is invertible.

Consider again the problem of finding a polynomial

$$f(t) = x_1 + x_2 t + x_3 t^2$$

that satisfies $f(1) = 8$, $f(2) = 5$, and $f(3) = -4$. The constraints yield the linear system

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ -4 \end{bmatrix} \tag{4.9}$$

Let's first find the inverse of the Vandermonde matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$$

using Lagrange polynomials. We calculate

$$f_1(t) = \frac{(t-2)(t-3)}{(1-2)(1-3)} = \frac{1}{2}\left(6 - 5t + t^2\right)$$

$$f_2(t) = \frac{(t-1)(t-3)}{(2-1)(2-3)} = -3 + 4t - t^2$$

$$f_3(t) = \frac{(t-1)(t-2)}{(3-1)(3-2)} = \frac{1}{2}\left(2 - 3t + t^2\right)$$

We next check that

$$A\begin{bmatrix} 3 \\ -2.5 \\ 0.5 \end{bmatrix} = \boldsymbol{e_1}, \qquad A\begin{bmatrix} -3 \\ 4 \\ -1 \end{bmatrix} = \boldsymbol{e_2}, \qquad A\begin{bmatrix} 1 \\ -1.5 \\ 0.5 \end{bmatrix} = \boldsymbol{e_3}$$

and thus

$$A^{-1} = \begin{bmatrix} 3 & -3 & 1 \\ -2.5 & 4 & -1.5 \\ 0.5 & -1 & 0.5 \end{bmatrix}$$

Finally we note that

$$\boldsymbol{x} = A^{-1}\boldsymbol{b} = \begin{bmatrix} 3 & -3 & 1 \\ -2.5 & 4 & -1.5 \\ 0.5 & -1 & 0.5 \end{bmatrix} \begin{bmatrix} 8 \\ 5 \\ -4 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ -3 \end{bmatrix}$$

and hence the unique interpolating polynomial is $f(t) = 5 + 6t - 3t^2$.

Exercise 4.26

Consider the problem of finding a polynomial

$$f(t) = x_1 + x_2 t + x_3 t^2 + x_4 t^3$$

such that $f(-2) = 3$, $f(-1) = -6$, $f(1) = 0$, and $f(3) = -2$. The constraints lead to the linear system $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 9 & 27 \end{bmatrix}$$

is a Vandermonde matrix. In this problem we will find A^{-1} using Lagrange polynomials. Let $c_1 = -2$, $c_2 = -1$, $c_3 = 1$, and $c_4 = 3$. Write each of the Lagrange polynomials f_1, f_2, f_3, f_4 given by

$$f_i(t) = \frac{(t - c_1)\cdots(t - c_{i-1})(t - c_{i+1})\cdots(t - c_n)}{(c_i - c_1)\cdots(c_i - c_{i-1})(c_i - c_{i+1})\cdots(c_i - c_n)}$$

in the form

$$f_i(t) = x_1 + x_2 t + x_3 t^2 + x_4 t^3$$

(you must determine the constants x_1, x_2, x_3, x_4 for each Lagrange polynomial). Use these constants to form A^{-1} and verify that $AA^{-1} = A^{-1}A = I_4$. Finally, use A^{-1} to solve $A\boldsymbol{x} = \boldsymbol{b}$ and find the unique interpolating polynomial $f(t)$ that fits the given points.

Let $c_1, c_2, \ldots, c_n$ be *distinct* scalars in $\mathbb{R}$ and let $f(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ for some scalars $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Suppose that $f(c_i) = 0$ for all i in $1, \ldots, n$.

$$\begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Since the Vandermonde matrix

$$A = \begin{bmatrix} 1 & c_1 & \cdots & c_1^{n-1} \\ 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \end{bmatrix}$$

is invertible, we have that

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A^{-1}\mathbf{0} = \mathbf{0}$$

and we have justified the following result.

Corollary 4.4.2

Let $c_1, c_2, \ldots, c_n$ be *distinct* scalars in $\mathbb{R}$ and let $f(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ for some scalars $x_1, x_2, \ldots, x_n \in \mathbb{R}$. If $f(c_i) = 0$ for all i in $1, \ldots, n$ then $x_i = 0$ for all i in $1, \ldots, n$ and thus $f(t) = \mathbf{0}$.

4.5. Polynomial Spaces

The set of polynomials

$$P(\mathbb{R}) = \operatorname{span}(\{1, t, t^2, \ldots\})$$

is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$. For $n \geq 0$

$$P_n(\mathbb{R}) = \operatorname{span}(\{1, t, t^2, \ldots, t^n\})$$

is a subspace of $P(\mathbb{R})$.

Let $V = P(\mathbb{R})$ and for any $n \geq 1$ consider the set $S = \{1, t, t^2, \ldots, t^{n-1}\}$. Suppose that

$$x_1 + x_2 t + \cdots + x_n t^{n-1} = \mathbf{0}$$

Let $f(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ and pick n distinct scalars $c_1, c_2, \ldots, c_n$ in $\mathbb{R}$. Then since $f(c_i) = 0$ for all i in $1, \ldots, n$ we must have $x_1 = x_2 = \cdots = x_n = 0$ by corollary 4.4.2. We have justified the following lemma.

Lemma 4.5.1

For any $n \geq 1$ the set $S = \{1, t, t^2, \ldots, t^{n-1}\}$ is a linearly independent set of n vectors in $P(\mathbb{R})$.

Let $V = P_3(\mathbb{R})$. Suppose that

$$\boldsymbol{f} = 3 + 3t + 2t^2 - t^3, \quad \boldsymbol{g}_1 = 1 + t + t^2 + t^3, \quad \boldsymbol{g}_2 = 1 + t + t^2, \quad \boldsymbol{g}_3 = 1 + t$$

We will check to see if $\boldsymbol{f} \in \operatorname{span}(\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\})$. Suppose that

$$3 + 3t + 2t^2 - t^3 = c_1(1 + t + t^2 + t^3) + c_2(1 + t + t^2) + c_3(1 + t)$$

for some $c_1, c_2, c_3 \in \mathbb{R}$. Collecting terms gives

$$(c_1 + c_2 + c_3 - 3)1 + (c_1 + c_2 + c_3 - 3)t + (c_1 + c_2 - 2)t^2 + (c_1 + 1)t^3 = \mathbf{0}$$

Since $\{1, t, t^2, t^3\}$ is a linearly independent set of 4 vectors by lemma 4.5.1, we must have

$$c_1 + c_2 + c_3 - 3 = 0, \qquad c_1 + c_2 - 2 = 0, \qquad c_1 + 1 = 0$$

We note that $c_1 = -1$, $c_2 = 2 - c_1 = 3$, $c_3 = 3 - c_1 - c_2 = 1$ works and hence $\boldsymbol{f} \in \operatorname{span}(\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\})$. Further note that since $\boldsymbol{f} \in \operatorname{span}(\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\})$ we have that $\{\boldsymbol{f}, \boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$ is a linearly dependent set.

Exercise 4.27

Is $S = \{1 + t + t^2, 3 + t^2, t + t^2\}$ a linearly independent set of 3 vectors in $P_2(\mathbb{R})$? Explain.

CHAPTER 5

Basis and Dimension

5.1. Finite-Dimensional Vector Spaces

Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a generating set for a vector space V containing n vectors. Then any vector $\boldsymbol{v} \in V$ can be written as

$$\boldsymbol{v} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n} \tag{5.1}$$

for some scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$. Suppose that

$$\boldsymbol{v} = d_1\boldsymbol{v_1} + d_2\boldsymbol{v_2} + \cdots + d_n\boldsymbol{v_n} \tag{5.2}$$

for some potentially *different* scalars $d_1, d_2, \ldots, d_n$. Subtracting equations (5.1) and (5.2) gives

$$(c_1 - d_1)\boldsymbol{v_1} + (c_2 - d_2)\boldsymbol{v_2} + \cdots + (c_n - d_n)\boldsymbol{v_n} = \boldsymbol{0} \tag{5.3}$$

We see from equation (5.3) that uniqueness of representation of the vector $\boldsymbol{v}$ is equivalent to linear independence of the generating set S.

Definition 5.1

Let V be a vector space. A set $\beta \subseteq V$ is a *basis* for V if
(1) S generates V.
(2) S is linearly independent.

Exercise 5.1

Let V be a vector space and let $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a subset of n vectors in V. Prove that β is a basis for V if and only only if each $\boldsymbol{v} \in V$ can be expressed as

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

for *unique* scalars $c_1, c_2, \ldots, c_n$.

The empty set $\varnothing$ is a basis for the vector space $V = \{\boldsymbol{0}\}$, the set $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is a basis for $\mathbb{R}^n$, and the set $\gamma = \{1, t, \ldots, t^n\}$ is a basis for $P_n(\mathbb{R})$.

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 & -4 \\ -1 & 1 & -4 & 2 \\ 1 & 2 & 1 & -5 \end{bmatrix}$$

Let $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\}$ where

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{a}_3 = \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \qquad \boldsymbol{a}_4 = \begin{bmatrix} -4 \\ 2 \\ -5 \end{bmatrix}$$

and note that $\operatorname{col}(A) = \operatorname{span}(S)$. Applying row operations to A results in the matrix

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

By inspecting the column vectors of C we see that $\boldsymbol{a}_3 = 3\boldsymbol{a}_1 - \boldsymbol{a}_2$. Thus $\text{col}(A) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_4\})$. We also note that $\boldsymbol{a}_4 = -3\boldsymbol{a}_1 - \boldsymbol{a}_2$. Thus $\text{col}(A) = \text{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2\})$. Finally suppose that

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 = \boldsymbol{0}$$

But then

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{c} = \boldsymbol{0}$$

and hence $\boldsymbol{c} = \boldsymbol{0}$. Thus $\beta = \{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is a linearly independent set and hence a basis for $\text{col}(A)$. Next we find a basis for $N(A)$. In equation form the system $C\boldsymbol{x} = \boldsymbol{0}$ reads

$$x_1 = -3x_3 + 3x_4, \qquad x_2 = x_3 + x_4$$

We set $t_1 = x_3$ and $t_2 = x_4$ and note that $\boldsymbol{x} \in N(A)$ if and only if

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3t_1 + 3t_2 \\ t_1 + t_2 \\ t_1 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Thus the set $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ generates $N(A)$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Next suppose that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 = \boldsymbol{0}$$

Then $c_1 = c_2 = 0$ and thus γ is linearly independent and hence a basis for $N(A)$.

Consider the matrix

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

Let $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ where

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 2 \end{bmatrix}, \qquad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \qquad \boldsymbol{a}_3 = \begin{bmatrix} -1 \\ 4 \\ 2 \\ -5 \end{bmatrix}$$

and note that $\text{col}(A) = \text{span}(S)$. Applying row operations to A results in the matrix

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

Suppose

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + c_3\boldsymbol{a}_3 = \boldsymbol{0}$$

Then

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{c} = \boldsymbol{0}$$

and hence $\boldsymbol{c} = \boldsymbol{0}$. Thus $\beta = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ is a linearly independent set of 3 vectors and hence a basis for $\text{col}(A)$. Since $N(A) = N(C) = \{\boldsymbol{0}\}$ we have that $\gamma = \varnothing$ is a basis for $N(A)$.

Exercise 5.2

Consider the matrix

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

Find a basis for $\text{col}(A)$ and $N(A)$.

Exercise 5.3

Consider the matrix

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

Find a basis for $\text{col}(A)$ and $N(A)$.

Definition 5.2

A vector space V with a *finite* basis is called *finite-dimensional*.

The vector spaces $\mathbb{R}^n$ and $P_n(\mathbb{R})$ are finite-dimensional. By theorem 4.4, the vector space $P(\mathbb{R})$ cannot have a finite basis β containing n vectors since β would be a generating set for $P(\mathbb{R})$ and $T = \{1, t, \ldots, t^n\}$ is a linearly independent set containing $n+1$ vectors in $P(\mathbb{R})$. Thus, $P(\mathbb{R})$ is an infinite-dimensional vector space.

Exercise 5.4

Let V be a finite-dimensional vector space. Prove that any basis β for V is a finite set.

Let V be a finite-dimensional vector space. By exercise 5.4 any basis for V is finite. Suppose

$$\beta = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}\}, \qquad \gamma = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n}\}$$

are bases for V containing m and n vectors respectively. Since β generates V and γ is linearly independent we have that $n \leq m$ by theorem 4.4. Since γ generates V and β is linearly independent we have that $m \leq n$ by theorem 4.4. Thus, $m = n$ and we have justified the following result.

Theorem 5.1

Let V be a finite-dimensional vector space. Then all bases for V are finite and have the same number of vectors.

Definition 5.3

Let V be a finite-dimensional vector space. The number of vectors in any basis for V is called the *dimension* of V and is denoted $\dim(V)$.

Note that $\dim(\mathbb{R}^n) = n$ and $\dim(P_n(\mathbb{R})) = n + 1$. The next theorem says that the dimension of a finite-dimensional vector space V is an upper bound on the size of a linearly independent set in V and a lower bound on the size of a generating set for V. We leave the proof as an exercise.

Theorem 5.2

Let V be a finite-dimensional vector space and let S be a set of m vectors in V.

(1) If S is linearly independent, then $m \leq \dim(V)$.
(2) If S generates V, then $m \geq \dim(V)$.

Exercise 5.5

Prove theorem 5.2.

5.2. Subspace Dimension

It turns out that any subspace W of a finite-dimensional vector space V must also be finite-dimensional but we will need a couple more results to prove it.

Lemma 5.2.1

Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a linearly independent set of k vectors in a vector space V. Let $\boldsymbol{v}$ be a vector in V that is *not* in S. Then $S \cup \{\boldsymbol{v}\}$ is linearly dependent if and only if $\boldsymbol{v} \in \operatorname{span}(S)$.

Proof. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a linearly independent set of k vectors in a vector space V. Let $\boldsymbol{v}$ be a vector in V that is *not* in S. If $\boldsymbol{v} \in \operatorname{span}(S)$, then $S \cup \{\boldsymbol{v}\}$ is linearly dependent by definition. Next, suppose $S \cup \{\boldsymbol{v}\}$ is linearly dependent. Since $\boldsymbol{v}$ is not in S, we must have $\boldsymbol{v} \in \operatorname{span}(S)$ or $\boldsymbol{v}_i \in \operatorname{span}((S \cup \{\boldsymbol{v}\}) - \{\boldsymbol{v}_i\})$ for some $1 \leq i \leq k$. Suppose the second case is true. Then

$$\boldsymbol{v}_i = c_1\boldsymbol{v}_1 + \cdots + c_{i-1}\boldsymbol{v}_{i-1} + c_{i+1}\boldsymbol{v}_{i+1} + \cdots + c_k\boldsymbol{v}_k + \alpha\boldsymbol{v}$$

where we must have $\alpha \neq 0$ since S is linearly independent. Solving for $\boldsymbol{v}$ gives

$$\boldsymbol{v} = \frac{1}{\alpha}\left(\boldsymbol{v}_i - (c_1\boldsymbol{v}_1 + \cdots + c_{i-1}\boldsymbol{v}_{i-1} + c_{i+1}\boldsymbol{v}_{i+1} + \cdots + c_k\boldsymbol{v}_k)\right)$$

and hence $\boldsymbol{v} \in \operatorname{span}(S)$. □

Lemma 5.2.2

Suppose a vector space V is not finite-dimensional. Then for all $n \geq 0$ there exists a linearly independent set T of n vectors.

Proof. Suppose V is not finite-dimensional. The proof is by mathematical induction on n. The induction begins with $n = 0$. In this case $T = \varnothing$ works since $\varnothing$ is linearly independent. Now suppose the thereom is true for some integer $n \geq 0$. We prove the theorem is true for $n + 1$. By the inductive hypothesis there exists a linearly independent set T of n vectors. Since V is not finite-dimensional we know that $\operatorname{span}(T) \neq V$ since otherwise T would be a finite basis. Pick $\boldsymbol{v} \in V$ so that $\boldsymbol{v} \notin \operatorname{span}(T)$. Since $\boldsymbol{v} \notin \operatorname{span}(T)$ we know that $T \cup \{\boldsymbol{v}\}$ is a linearly independent set of $n+1$ vectors by the contrapositive of lemma 5.2.1. □

Suppose that W is a subspace of a finite-dimensional vector space V and let β be a basis for V containing $\dim(V)$ vectors. By theorem 5.2 the subspace $W \subseteq V$ cannot contain a linearly independent set of more than $\dim(V)$ vectors. Thus, by lemma 5.2.2 we must have that W is also finite-dimensional and furthermore $\dim(W) \leq \dim(V)$. We have justified the next theorem.

Theorem 5.3

A subspace W of a finite-dimensional vector space V is finite-dimensional and $\dim(W) \leq \dim(V)$.

Let $A \in \mathbb{R}^{m\times n}$. By theorem 5.3 we have that $\text{col}(A)$ and $N(A)$ are finite-dimensional and furthermore

$$\dim(\text{col}(A)) \leq m, \qquad \dim(N(A)) \leq n$$

since $\text{col}(A)$ is a subspace of $\mathbb{R}^m$ and $N(A)$ is a subspace of $\mathbb{R}^n$.

Definition 5.4

Let $A \in \mathbb{R}^{m\times n}$. We define

$$\text{rank}(A) = \dim(\text{col}(A)), \qquad \text{nullity}(A) = \dim(N(A))$$

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 & -4 \\ -1 & 1 & -4 & 2 \\ 1 & 2 & 1 & -5 \end{bmatrix}$$

We have shown previously that $\beta = \{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is a basis for $\text{col}(A)$ where $\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, $\boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a basis for $N(A)$ where $\boldsymbol{v}_1 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$. Thus we have that $\text{rank}(A) = 2$ and $\text{nullity}(A) = 2$.

Next suppose

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

We have shown that $\beta = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ is a basis for $\text{col}(A)$ where $\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 2 \end{bmatrix}$, $\boldsymbol{a}_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \end{bmatrix}$, $\boldsymbol{a}_3 = \begin{bmatrix} -1 \\ 4 \\ 2 \\ -5 \end{bmatrix}$ and $\gamma = \varnothing$ is a basis for $N(A)$. Thus we have that $\text{rank}(A) = 3$ and $\text{nullity}(A) = 0$.

Exercise 5.6

Consider the matrix

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

Calculate $\text{rank}(A)$ and $\text{nullity}(A)$.

Exercise 5.7

Consider the matrix

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

Calculate $\text{rank}(A)$ and $\text{nullity}(A)$.

Exercise 5.8

Let V be a finite-dimensional vector space and supppose that U and W are subspaces of V.

(1) Prove that if $V = U + W$, then $\dim(V) \leq \dim(U) + \dim(W)$.
(2) Prove that if $V = U \oplus W$, then $\dim(V) = \dim(U) + \dim(W)$.

Suppose $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_n}$. Let $\beta = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ be a basis for $\mathrm{col}(A)$ containing $k = \mathrm{rank}(A)$ vectors. Let $W \in \mathbb{R}^{m\times k}$ be the matrix with column vectors $\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}$. Since β generates $\mathrm{col}(A)$, for all i in $1, \ldots, n$ we can write

$$\boldsymbol{a_i} = b_{1i}\boldsymbol{w_1} + b_{2i}\boldsymbol{w_2} + \cdots + b_{ki}\boldsymbol{w_k}$$

for some scalars $b_{1i}, b_{2i}, \ldots, b_{ki} \in \mathbb{R}$. In matrix form we have

$$A = WB$$

where $B \in \mathbb{R}^{k\times n}$ is the matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kn} \end{bmatrix}$$

Using equation (1.12) we have

$$A^T = B^T W^T \tag{5.4}$$

where B^T has column vectors $\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_k}$. Equation (5.4) shows that the column vectors of A^T are in $\mathrm{span}(\gamma)$ where $\gamma = \{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_k}\}$ and corollary 4.1.1 implies that $\mathrm{col}(A^T)$ is generated by γ. By theorem 5.2 we thus have that

$$\mathrm{rank}(A^T) \leq \mathrm{rank}(A) \tag{5.5}$$

Plugging A^T into equation (5.5) gives

$$\mathrm{rank}(A) \leq \mathrm{rank}(A^T)$$

and we have justified the following theorem.

Theorem 5.4

Let $A \in \mathbb{R}^{m\times n}$. Then

$$\mathrm{rank}(A) = \mathrm{rank}(A^T)$$

Exercise 5.9

Let $A \in \mathbb{R}^{m\times n}$. Prove that $\mathrm{rank}(A) \leq m$ and $\mathrm{rank}(A) \leq n$.

5.3. Reduction and Extension

The next result gives an algorithm for reducing a finite generating set to a finite basis. This result implies that a vector space that can be generated by a finite set must be finite-dimensional.

Lemma 5.3.1

Let V be a vector space with finite generating set $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$. Define $S_0 = \varnothing$ and for k in $1, \dots, n$ define $S_k = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}\}$. Define $T_0 = \varnothing$ and for k in $0, \dots, n-1$ define $T_{k+1} = T_k$ if $\boldsymbol{v_{k+1}} \in \text{span}(T_k)$ or $T_{k+1} = T_k \cup \{\boldsymbol{v_{k+1}}\}$ otherwise. Then for all k in $0, \dots, n$ we have that

(1) T_k is linearly independent.
(2) $\text{span}(S_k) = \text{span}(T_k)$.

Proof. The proof is by mathematical induction on k. The induction begins with $k = 0$. In this case $T_0 = \varnothing$ is linearly independent and since $S_0 = T_0$ we have $\text{span}(S_0) = \text{span}(T_0)$. Now suppose the the lemma is true for some integer k in $0, \dots, n-1$. We prove the lemma is true for $k+1$. By the inductive hypothesis we have T_k is linearly independent and $\text{span}(S_k) = \text{span}(T_k)$. Suppose that $\boldsymbol{v_{k+1}} \in \text{span}(T_k)$. Then $T_{k+1} = T_k$ is linearly independent and

$$\text{span}(S_{k+1}) = \text{span}(S_k) = \text{span}(T_k) = \text{span}(T_{k+1})$$

by lemma 4.3.1 since $\boldsymbol{v_{k+1}} \in \text{span}(S_k)$. Suppose that $\boldsymbol{v_{k+1}} \notin \text{span}(T_k)$. Then $T_{k+1} = T_k \cup \{\boldsymbol{v_{k+1}}\}$ is linearly independent by the contrapositive of lemma 5.2.1. Since

$$S_k \subseteq \text{span}(S_k) = \text{span}(T_k) \subseteq \text{span}(T_{k+1})$$

by exercise 4.5 and

$$\boldsymbol{v_{k+1}} \in T_{k+1} \subseteq \text{span}(T_{k+1})$$

we have that $S_{k+1} \subseteq \text{span}(T_{k+1})$ and hence $\text{span}(S_{k+1}) \subseteq \text{span}(T_{k+1})$ by corollary 4.1.1. Finally, since $T_{k+1} \subseteq S_{k+1}$ we have $\text{span}(T_{k+1}) \subseteq \text{span}(S_{k+1})$ by exercise 4.5. □

The next theorem follows from lemma 5.3.1. We leave the proof as an exercise.

Theorem 5.5

Let V be a vector space and S be a finite generating set for V. Then there exists a set $T \subseteq S$ such that T is a basis for V and hence V is finite-dimensional.

Consider the matrix

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

with column vectors $\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}, \boldsymbol{a_4}$. Applying the reduction algorithm to find a basis for $\text{col}(A)$ gives

$$T_0 = \varnothing, \qquad T_1 = \{\boldsymbol{a_1}\}, \qquad T_2 = \{\boldsymbol{a_1}, \boldsymbol{a_2}\}, \qquad T_3 = \{\boldsymbol{a_1}, \boldsymbol{a_2}\}, \qquad T_4 = \{\boldsymbol{a_1}, \boldsymbol{a_2}\}$$

Thus $T = T_4 = \{\boldsymbol{a_1}, \boldsymbol{a_2}\}$ is a basis for $\text{col}(A)$ and $\text{rank}(A) = 2$.

Exercise 5.10

Consider the matrix

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

Use the reduction algorithm to find a basis for $\text{col}(A)$ and calculate $\text{rank}(A)$.

Let V be a finite-dimensional vector space and let S be a generating set for V containing $m = \dim(V)$ vectors. By theorem 5.5, there exists a basis T for V such that $T \subseteq S$. But by theorem 5.1, T must have $\dim(V)$ vectors and hence $S = T$. We have justified the following result.

Corollary 5.3.1

Let V be a finite-dimensional vector space and let S be a subset of V containing $m = \dim(V)$ vectors. If S generates V, then S is a basis for V.

Exercise 5.11

Let $A \in \mathbb{R}^{m\times n}$. Prove $\text{nullity}(A) = 0$ if and only if $\text{rank}(A) = n$.

Exercise 5.12

Let V be a finite-dimensional vector space and supppose that U and W are subspaces of V. Prove that if $V = U + W$ and $\dim(V) = \dim(U) + \dim(W)$, then $V = U \oplus W$.

The next result gives an algorithm for extending a linearly independent set in a finite-dimensional vector space V to a basis for V. We leave the proof as an exercse.

Lemma 5.3.2

Let V be a finite-dimensional vector space and let S be a linearly independent set in V. Let $\beta = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ be a basis for V containing $n = \dim(V)$ vectors. Define $S_0 = S$ and for k in $1, \ldots, n$ define $S_k = S \cup \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$. Define $T_0 = S$ and for k in $0, \ldots, n-1$ define $T_{k+1} = T_k$ if $\boldsymbol{v_{k+1}} \in \text{span}(T_k)$ or $T_{k+1} = T_k \cup \{\boldsymbol{v_{k+1}}\}$ otherwise. Then for all k in $0, \ldots, n$ we have that

(1) T_k is linearly independent.
(2) $\text{span}(S_k) = \text{span}(T_k)$.

Exercise 5.13

Prove lemma 5.3.2.

The next theorem follows from lemma 5.3.2. We leave the proof as an exercise.

Theorem 5.6

Let V be a finite-dimensional vector space and suppose S is a linearly independent subset of V. Then there exists a basis T for V such that $S \subseteq T$.

Consider the linearly independent set $S = \{\boldsymbol{v}\}$ in $\mathbb{R}^3$ where $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Applying the extension algorithm with the basis $\beta = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ to find a basis for $\mathbb{R}^3$ containing S gives

$$T_0 = \{\boldsymbol{v}\}, \qquad T_1 = \{\boldsymbol{v}, \boldsymbol{e_1}\}, \qquad T_2 = \{\boldsymbol{v}, \boldsymbol{e_1}\}, \qquad T_3 = \{\boldsymbol{v}, \boldsymbol{e_1}, \boldsymbol{e_3}\}$$

Thus $T = T_3 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$ containing the set S.

Exercise 5.14

Consider the linearly independent set $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ in $\mathbb{R}^4$. Use the extension algorithm with the basis $\beta = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \boldsymbol{e_4}\}$ to find a basis for $\mathbb{R}^4$ containing S.

Let V be a finite-dimensional vector space and let S be a linearly independent set in V containing $m = \dim(V)$ vectors. By theorem 5.6, there exists a basis T for V such that $S \subseteq T$. But by theorem 5.1, T must have $\dim(V)$ vectors and hence $S = T$. We have justified the following result.

Corollary 5.3.2

Let V be a finite-dimensional vector space and let S be a set in V containing $m = \dim(V)$ vectors. If S is linearly independent, then S is a basis for V.

Exercise 5.15

Which (if any) of the following sets are a basis for $\mathbb{R}^3$?

$$S = \left\{ \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ -1 \end{bmatrix} \right\}, \qquad T = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -4 \\ 3 \end{bmatrix} \right\}$$

Exercise 5.16

Is the set $S = \{1 + 2t + t^2, 3 + 6t + t^2, t + t^2\}$ a basis for $P_2(\mathbb{R})$? Explain.

Exercise 5.17

Let $c_1, c_2, \ldots, c_n$ be *distinct* scalars in $\mathbb{R}$. Recall that the Lagrange polynomials $\boldsymbol{f_1}, \boldsymbol{f_2}, \ldots, \boldsymbol{f_n}$ are defined by

$$f_i(t) = \frac{(t - c_1) \cdots (t - c_{i-1})(t - c_{i+1}) \cdots (t - c_n)}{(c_i - c_1) \cdots (c_i - c_{i-1})(c_i - c_{i+1}) \cdots (c_i - c_n)}$$

Prove that $\beta = \{\boldsymbol{f_1}, \boldsymbol{f_2}, \ldots, \boldsymbol{f_n}\}$ is a basis for $P_{n-1}(\mathbb{R})$.

Let W be a subspace of a finite-dimensional vector space V. If $W = V$, then $\dim(W) = \dim(V)$. Next suppose that $\dim(W) = \dim(V)$. Let β be a basis for W with $\dim(W) = \dim(V)$ vectors. Since $W \subseteq V$, β is a linearly independent subset of V with $\dim(V)$ vectors and is hence a basis for V by corollary 5.3.2. Thus we have $W = \operatorname{span}(\beta) = V$. We have justified the following theorem.

Theorem 5.7

Let W be a subspace of a finite-dimensional vector space V. Then we have that $W = V$ if and only if $\dim(W) = \dim(V)$.

Exercise 5.18

Let $A \in \mathbb{R}^{m \times n}$. Prove that $\operatorname{col}(A) = \mathbb{R}^m$ if and only if $\operatorname{rank}(A) = m$.

Exercise 5.19

Let $A \in \mathbb{R}^{n \times n}$. Prove that the following statements are equivalent.

(1) $\operatorname{nullity}(A) = 0$.
(2) $\operatorname{rank}(A) = n$.
(3) A is invertible.

CHAPTER 6

Linear Maps

6.1. Definition and Examples

Definition 6.1

Let V and W be vector spaces. The map $T : V \to W$ is called *linear* if for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $a, b \in \mathbb{R}$ we have that

$$T(a\boldsymbol{x} + b\boldsymbol{y}) = aT\boldsymbol{x} + bT\boldsymbol{y} \tag{6.1}$$

Let V and W be vector spaces and $T : V \to W$ be a linear map. We note that

$$\begin{aligned} T\mathbf{0} + \mathbf{0} &= T\mathbf{0} && \text{Axiom (VS 3)} \\ &= T(\mathbf{0} + \mathbf{0}) && \text{Axiom (VS 3)} \\ &= T(1\mathbf{0} + 1\mathbf{0}) && \text{Axiom (VS 5)} \\ &= 1T\mathbf{0} + 1T\mathbf{0} && \text{Equation (6.1)} \\ &= T\mathbf{0} + T\mathbf{0} && \text{Axiom (VS 5)} \end{aligned}$$

and thus we have

$$T\mathbf{0} = \mathbf{0}$$

by theoreom 3.1. We have justified the following result.

Lemma 6.1.1

Let V and W be vector spaces and $T : V \to W$ be a linear map. Then

$$T\mathbf{0} = \mathbf{0}$$

Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 1 \\ x_2 \end{bmatrix}$. Since $T \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, T is not linear by lemma 6.1.1.

Let $A \in \mathbb{R}^{m \times n}$ and define the map $T : \mathbb{R}^n \to \mathbb{R}^m$ for $\boldsymbol{x} \in \mathbb{R}^n$ using the formula

$$T\boldsymbol{x} = A\boldsymbol{x}$$

Using equation (1.3) we note that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$ we have that

$$T(a\boldsymbol{x} + b\boldsymbol{y}) = A(a\boldsymbol{x} + b\boldsymbol{y}) = aA\boldsymbol{x} + bA\boldsymbol{y} = aT\boldsymbol{x} + bT\boldsymbol{y}$$

and thus T is a linear map. We have justified the following result.

Lemma 6.1.2

If $A \in \mathbb{R}^{m \times n}$, then the map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined for $\boldsymbol{x} \in \mathbb{R}^n$ using the formula

$$T\boldsymbol{x} = A\boldsymbol{x}$$

is linear.

Consider the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined using the formula

$$T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}$$

Since $T\boldsymbol{x} = A\boldsymbol{x}$ where $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, the map T is linear by lemma 6.1.2.

Exercise 6.1

Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined using the formula

$$T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - 2x_2 + x_3 \\ x_2 - 5x_3 \end{bmatrix}$$

Is T a linear map? Explain.

Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and consider the map $T : V \to V$ defined by $(T\boldsymbol{f})(t) = (t^2 + t)f(t)$. Let $\boldsymbol{f}, \boldsymbol{g} \in V$ and $a, b \in \mathbb{R}$. Note that for all $t \in \mathbb{R}$ we have that

$$\begin{aligned}
T(a\boldsymbol{f} + b\boldsymbol{g})(t) &= (t^2 + t)(a\boldsymbol{f} + b\boldsymbol{g})(t) \\
&= (t^2 + t)((a\boldsymbol{f})(t) + (b\boldsymbol{g})(t)) \\
&= (t^2 + t)(a\boldsymbol{f}(t) + b\boldsymbol{g}(t)) \\
&= a(t^2 + t)\boldsymbol{f}(t) + b(t^2 + t)\boldsymbol{g}(t) \\
&= a(T\boldsymbol{f})(t) + b(T\boldsymbol{g})(t) \\
&= (aT\boldsymbol{f})(t) + (bT\boldsymbol{g})(t) \\
&= (aT\boldsymbol{f} + bT\boldsymbol{g})(t)
\end{aligned}$$

Thus we have that

$$T(a\boldsymbol{f} + b\boldsymbol{g}) = aT\boldsymbol{f} + bT\boldsymbol{g}$$

and hence T is a linear map.

Exercise 6.2

Consider the map $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ defined for $\boldsymbol{f} \in P_2(\mathbb{R})$ by

$$(T\boldsymbol{f})(t) = tf(t) + f'(t)$$

Prove that the map T is linear.

Exercise 6.3

Let V and W be vector spaces and $T : V \to W$ and $U : V \to W$ be linear maps. Let $a \in \mathbb{R}$. The map $T + aU$ is defined for $\boldsymbol{x} \in V$ by

$$(T + aU)\boldsymbol{x} = T\boldsymbol{x} + aU\boldsymbol{x}$$

Prove that the map $T + aU$ is linear.

Exercise 6.4

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be vectors in a vector space V. Define the map $T : \mathbb{R}^n \to V$ by

$$T\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n$$

Prove that the map T is linear.

Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map and that

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

Using equation (6.1) we calculate

$$T\begin{bmatrix}-1\\2\end{bmatrix} = T\left(-1\begin{bmatrix}1\\0\end{bmatrix}+2\begin{bmatrix}0\\1\end{bmatrix}\right) = -1T\begin{bmatrix}1\\0\end{bmatrix}+2T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\-3\end{bmatrix}+\begin{bmatrix}4\\0\end{bmatrix} = \begin{bmatrix}5\\-3\end{bmatrix}$$

We again use equation (6.1) and note that for any $\boldsymbol{x} \in \mathbb{R}^2$ we have

$$T\boldsymbol{x} = T\begin{bmatrix}x_1\\x_2\end{bmatrix} = T\left(x_1\begin{bmatrix}1\\0\end{bmatrix}+x_2\begin{bmatrix}0\\1\end{bmatrix}\right) = x_1\begin{bmatrix}-1\\3\end{bmatrix}+x_2\begin{bmatrix}2\\0\end{bmatrix} = A\begin{bmatrix}x_1\\x_2\end{bmatrix} = A\boldsymbol{x}$$

where $A = \begin{bmatrix}-1 & 2\\3 & 0\end{bmatrix}$. Again we see that

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

Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear map and that

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

Suppose we want to know $T\begin{bmatrix}8\\11\end{bmatrix}$. We look for $\boldsymbol{x} = \begin{bmatrix}x_1\\x_2\end{bmatrix}$ such that

$$x_1\begin{bmatrix}1\\1\end{bmatrix}+x_2\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}1 & 2\\1 & 3\end{bmatrix}\boldsymbol{x} = \begin{bmatrix}8\\11\end{bmatrix}$$

It is easy to check that the matrix $\begin{bmatrix}1 & 2\\1 & 3\end{bmatrix}$ is invertible and $\begin{bmatrix}1 & 2\\1 & 3\end{bmatrix}^{-1} = \begin{bmatrix}3 & -2\\-1 & 1\end{bmatrix}$. Thus we have

$$\boldsymbol{x} = \begin{bmatrix}3 & -2\\-1 & 1\end{bmatrix}\begin{bmatrix}8\\11\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$$

and furthermore

$$T\begin{bmatrix}8\\11\end{bmatrix} = T\left(2\begin{bmatrix}1\\1\end{bmatrix}+3\begin{bmatrix}2\\3\end{bmatrix}\right) = 2T\begin{bmatrix}1\\1\end{bmatrix}+3T\begin{bmatrix}2\\3\end{bmatrix} = 2\begin{bmatrix}1\\0\\2\end{bmatrix}+3\begin{bmatrix}1\\-1\\4\end{bmatrix} = \begin{bmatrix}5\\-3\\16\end{bmatrix}$$

Next we note that since

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

we have that

$$T\begin{bmatrix}1\\0\end{bmatrix} = T\left(3\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}2\\3\end{bmatrix}\right) = 3\begin{bmatrix}1\\0\\2\end{bmatrix}-\begin{bmatrix}1\\-1\\4\end{bmatrix} = \begin{bmatrix}2\\1\\2\end{bmatrix}$$

$$T\begin{bmatrix}0\\1\end{bmatrix} = T\left(-2\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}2\\3\end{bmatrix}\right) = -2\begin{bmatrix}1\\0\\2\end{bmatrix}+\begin{bmatrix}1\\-1\\4\end{bmatrix} = \begin{bmatrix}-1\\-1\\0\end{bmatrix}$$

and hence $T\boldsymbol{x} = A\boldsymbol{x}$ where

$$A = \left[T\begin{bmatrix}1\\0\end{bmatrix} \quad T\begin{bmatrix}0\\1\end{bmatrix}\right] = \begin{bmatrix}2 & -1\\1 & -1\\2 & 0\end{bmatrix}$$

Finally, we verify that

$$A\begin{bmatrix}8\\11\end{bmatrix} = \begin{bmatrix}2 & -1\\1 & -1\\2 & 0\end{bmatrix}\begin{bmatrix}8\\11\end{bmatrix} = \begin{bmatrix}5\\-3\\16\end{bmatrix} = T\begin{bmatrix}8\\11\end{bmatrix}$$

Exercise 6.5

Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map and that

$$T\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}2\\-3\\4\end{bmatrix}, \qquad T\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}3\\0\\1\end{bmatrix}, \qquad T\begin{bmatrix}1\\2\\-1\end{bmatrix} = \begin{bmatrix}5\\1\\0\end{bmatrix}$$

(1) Calculate $T\begin{bmatrix}2\\1\\1\end{bmatrix}$.

(2) Find a matrix $A \in \mathbb{R}^{3\times 3}$ such that for all $\boldsymbol{x} \in \mathbb{R}^3$ we have that $T\boldsymbol{x} = A\boldsymbol{x}$.

Exercise 6.6

Let V and W be vector spaces and suppose the map $T : V \to W$ is linear. Use induction on $k \geq 2$ to show that for all $c_1, c_2, \ldots, c_k \in \mathbb{R}$ and all $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in V$ we have

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = c_1T\boldsymbol{v}_1 + c_2T\boldsymbol{v}_2 + \cdots + c_kT\boldsymbol{v}_k \tag{6.2}$$

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Using equation (6.2) we note that for any $\boldsymbol{x} \in \mathbb{R}^n$

$$T\boldsymbol{x} = T\left(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n\right) = x_1T\boldsymbol{e}_1 + x_2T\boldsymbol{e}_2 + \cdots + x_nT\boldsymbol{e}_n = A\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = A\boldsymbol{x}$$

where $A \in \mathbb{R}^{m\times n}$ is the matrix $A = \begin{bmatrix}T\boldsymbol{e}_1 & T\boldsymbol{e}_2 & \cdots & T\boldsymbol{e}_n\end{bmatrix}$. We have justified the following result.

Lemma 6.1.3

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then there exists $A \in \mathbb{R}^{m\times n}$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$T\boldsymbol{x} = A\boldsymbol{x}$$

Exercise 6.7

Suppose that V and W are vector spaces and that $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is a generating set for V. Prove that if $T : V \to W$ and $U : V \to W$ are linear maps such that $T\boldsymbol{v}_i = U\boldsymbol{v}_i$ for all $1 \leq i \leq n$, then $T = U$.

6.2. Null Space and Range

Definition 6.2

Let V and W be vector spaces and let $T : V \to W$ be a linear map. The map T is called *one-to-one* if for all vectors $\boldsymbol{x}, \boldsymbol{y} \in V$ we have $T\boldsymbol{x} = T\boldsymbol{y}$ only if $\boldsymbol{x} = \boldsymbol{y}$.

Let V and W be vector spaces and let $T : V \to W$ be a linear map. Suppose $T\boldsymbol{x} = T\boldsymbol{y}$. Then since T is linear we have that

$$T(\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{0} \tag{6.3}$$

We see from equation (6.3) that equality of $\boldsymbol{x}$ and $\boldsymbol{y}$ depends on which vectors T maps to $\boldsymbol{0}$.

Definition 6.3

Let V and W be vector spaces and let $T : V \to W$ be a linear map. The *null space* of T, denoted $N(T)$, is defined as

$$N(T) = \{\boldsymbol{x} \in V : T\boldsymbol{x} = \mathbf{0}\}$$

Exercise 6.8

Let V and W be vector spaces and let $T : V \to W$ be a linear map. Prove that T is one-to-one if and only $N(T) = \{\mathbf{0}\}$.

Let V and W be vector spaces and let $T : V \to W$ be a linear map. We make two observations.

(1) Since $T\mathbf{0} = \mathbf{0}$ by lemma 6.1.1, we have that $\mathbf{0} \in N(T)$.

(2) Suppose $\boldsymbol{x}, \boldsymbol{y} \in N(T)$ and let $a, b \in \mathbb{R}$. Then we have $T\boldsymbol{x} = \mathbf{0}$ and $T\boldsymbol{y} = \mathbf{0}$. Since T is a linear map we have

$$T(a\boldsymbol{x} + b\boldsymbol{y}) = aT\boldsymbol{x} + bT\boldsymbol{y} = a\mathbf{0} + b\mathbf{0} = \mathbf{0}$$

and $a\boldsymbol{x} + b\boldsymbol{y}$ is in $N(T)$.

Thus by theorem 3.2 we have the following result.

Lemma 6.2.1

Let V and W be vector spaces and let $T : V \to W$ be linear. Then $N(T)$ is a subspace of V.

Definition 6.4

Let V and W be vector spaces and let $T : V \to W$ be a linear map. If $N(T)$ is finite-dimensional, we define

$$\text{nullity}(T) = \dim(N(T))$$

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that V is finite-dimensional. Since $N(T)$ is a subspace of V we have that $N(T)$ is finite-dimensional and furthermore

$$\text{nullity}(T) \leq \dim(V) \tag{6.4}$$

by theorem 5.3.

Exercise 6.9

Let $A \in \mathbb{R}^{m\times n}$ and define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T\boldsymbol{x} = A\boldsymbol{x}$. Prove that $N(T) = N(A)$ and hence $\text{nullity}(T) = \text{nullity}(A)$.

Definition 6.5

Let V and W be vector spaces and let $T : V \to W$ be linear. Let $S \subseteq V$ be a subset of V. We define

$$T(S) = \{T\boldsymbol{v} : \boldsymbol{v} \in S\}$$

Let V and W be vector spaces and let $T : V \to W$ be a linear map that is one-to-one. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a linearly independent set of k vectors in V. Suppose that

$$c_1T\boldsymbol{v}_1 + c_2T\boldsymbol{v}_2 + \cdots + c_kT\boldsymbol{v}_k = \mathbf{0}$$

Since T is linear, we have that

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = \mathbf{0}$$

Since T is one-to-one, we have that $N(T) = \{\mathbf{0}\}$ by exercise 6.8 and hence

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$$

But then since S is linearly independent, we have that $c_1 = c_2 = \cdots = c_k = 0$ and hence the set

$$T(S) = \{T\boldsymbol{v}_1, T\boldsymbol{v}_2, \ldots, T\boldsymbol{v}_k\}$$

is a linearly independent set of k vectors in W. We have justified the following result.

Lemma 6.2.2

Let V and W be vector spaces and let $T : V \to W$ be a linear map that is one-to-one. Suppose that $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is a linearly independent set of k vectors in V. Then $T(S) = \{T\boldsymbol{v}_1, T\boldsymbol{v}_2, \ldots, T\boldsymbol{v}_k\}$ is a linearly independent set of k vectors in W.

Definition 6.6

Let V and W be vector spaces and let $T : V \to W$ be a linear map. The *range* of T, denoted $R(T)$, is defined as

$$R(T) = T(V) = \{T\boldsymbol{v} : \boldsymbol{v} \in V\}$$

Definition 6.7

Let V and W be vector spaces and let $T : V \to W$ be a linear map. The map T is called *onto* if $R(T) = W$.

Exercise 6.10

Let V and W be vector spaces and let $T : V \to W$ be a linear map. Prove that T is onto if and only if for any vector $\boldsymbol{w} \in W$ there exists a vector $\boldsymbol{v} \in V$ such that $T\boldsymbol{v} = \boldsymbol{w}$.

Exercise 6.11

Suppose $A \in \mathbb{R}^{m\times n}$ and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear map defined by $T\boldsymbol{x} = A\boldsymbol{x}$.

(1) Prove that the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for *all* $\boldsymbol{b} \in \mathbb{R}^m$ if and only if T is onto.
(2) Prove that the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a *unique* solution for *all* $\boldsymbol{b} \in \operatorname{col}(A)$ if and only if T is one-to-one.
(3) Prove that the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a *unique* solution for *all* $\boldsymbol{b} \in \mathbb{R}^m$ if and only if T is one-to-one and onto.

Let V and W be vector spaces and let $T : V \to W$ be a linear map. We make two observations.

(1) Since $T\mathbf{0} = \mathbf{0}$ by lemma 6.1.1, we have that $\mathbf{0} \in R(T)$.
(2) Suppose $\boldsymbol{x}, \boldsymbol{y} \in R(T)$ and let $a, b \in \mathbb{R}$. Then we have $T\boldsymbol{u} = \boldsymbol{x}$ and $T\boldsymbol{v} = \boldsymbol{y}$ for some vectors $\boldsymbol{u}, \boldsymbol{v} \in V$. Since T is a linear map we have

$$T(a\boldsymbol{u} + b\boldsymbol{v}) = aT\boldsymbol{u} + bT\boldsymbol{v} = a\boldsymbol{x} + b\boldsymbol{y}$$

and $a\boldsymbol{x} + b\boldsymbol{y}$ is in $R(T)$.

Thus by theorem 3.2 we have the following result.

Lemma 6.2.3

Let V and W be vector spaces and let $T : V \to W$ be linear. Then $R(T)$ is a subspace of W.

Definition 6.8

Let V and W be vector spaces and let $T : V \to W$ be a linear map. If $R(T)$ is finite-dimensional, we define

$$\operatorname{rank}(T) = \dim(R(T))$$

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that W is finite-dimensional. Since $R(T)$ is a subspace of W we have that $R(T)$ is finite-dimensional and furthermore

$$\operatorname{rank}(T) \leq \dim(W) \tag{6.5}$$

by theorem 5.3.

Exercise 6.12

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that W is finite-dimensional. Prove that T is onto if and only if

$$\operatorname{rank}(T) = \dim(W)$$

Let V and W be vector spaces and let $T : V \to W$ be linear. Let $S \subseteq V$ be a generating set for V. Since $T(S) \subseteq T(V) = R(T)$ and $R(T)$ is a subspace of W we have that $\operatorname{span}(T(S)) \subseteq R(T)$ by theorem 4.1. Next, let $y \in R(T)$. Then $T\boldsymbol{x} = \boldsymbol{y}$ for some $\boldsymbol{x} \in V$. Since S generates V we have that

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$$

for some $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in V$ and $c_1, c_2, \ldots, c_k \in \mathbb{R}$. We note that

$$\boldsymbol{y} = T\boldsymbol{x} = T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = c_1T\boldsymbol{v}_1 + c_2T\boldsymbol{v}_2 + \cdots + c_kT\boldsymbol{v}_k \in \operatorname{span}(T(S))$$

We have justified the following result.

Lemma 6.2.4

Let V and W be vector spaces and let $T : V \to W$ be linear. If $S \subseteq V$ is a generating set for V, then

$$R(T) = \operatorname{span}(T(S))$$

Suppose $A \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T\boldsymbol{x} = A\boldsymbol{x}$. Since the set $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ generates $\mathbb{R}^n$ we have by lemma 6.2.4 that

$$\begin{aligned}
R(T) &= \operatorname{span}(T(S)) \\
&= \operatorname{span}(\{T\boldsymbol{e}_1, T\boldsymbol{e}_2, \ldots, T\boldsymbol{e}_n\}) \\
&= \operatorname{span}(\{A\boldsymbol{e}_1, A\boldsymbol{e}_2, \ldots, A\boldsymbol{e}_n\}) \\
&= \operatorname{span}(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}) \\
&= \operatorname{col}(A)
\end{aligned}$$

We have justified the following result.

Lemma 6.2.5

Suppose $A \in \mathbb{R}^{m\times n}$ and define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T\boldsymbol{x} = A\boldsymbol{x}$. Then $R(T) = \operatorname{col}(A)$ and hence $\operatorname{rank}(T) = \operatorname{rank}(A)$.

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that V is finite-dimensional. Let β be a basis for V. By lemma 6.2.4, the finite set $T(\beta)$ generates $R(T)$. Thus $R(T)$ is finite-dimensional by theorem 5.5. Furthermore, since $T(\beta)$ has at most $\dim(V)$ vectors we have that

$$\operatorname{rank}(T) \leq \dim(V) \tag{6.6}$$

by theorem 5.2.

Exercise 6.13

Let V and W be vector spaces. Suppose V_1 is a subspace of V, W_1 is a subspace of W, and $T : V \to W$ is a linear map. Show that $U_1 = \{\boldsymbol{x} \in V : T\boldsymbol{x} \in W_1\}$ is a subspace of V and that $U_2 = \{T\boldsymbol{x} : \boldsymbol{x} \in V_1\}$ is a subspace of W.

Exercise 6.14

Let V and W be finite-dimensional vector spaces and $T : V \to W$ be a linear map.

(1) Prove that if T is one-to-one, then $\dim(V) \leq \dim(W)$.
(2) Prove that if T is onto, then $\dim(V) \geq \dim(W)$.
(3) Conclude that if T is one-to-one and onto, then $\dim(V) = \dim(W)$.

Consider the map $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ defined for $\boldsymbol{f} \in P_2(\mathbb{R})$ by

$$(T\boldsymbol{f})(t) = tf(t) + f'(t)$$

By exercise 6.2, we have that T is a linear map. Let $a + bt + ct^2$ be an arbitrary vector in $P_2(\mathbb{R})$ and suppose that

$$T(a + bt + ct^2) = \mathbf{0}$$

Then we have that

$$\mathbf{0} = t(a + bt + ct^2) + (b + 2ct) = b + (a + 2c)t + bt^2 + ct^3$$

Since the set $\{1, t, t^2, t^3\}$ is linearly independent, we must have

$$b = 0, \qquad a + 2c = 0, \qquad c = 0$$

which implies that $a = b = c = 0$. We have shown that $T\boldsymbol{f} = \mathbf{0}$ if and only if $\boldsymbol{f} = \mathbf{0}$ and thus $N(T) = \{\mathbf{0}\}$, $\text{nullity}(T) = 0$, and T is one-to-one by exercise 6.8. Since $S = \{1, t, t^2\}$ is a basis for V we have that

$$R(T) = \text{span}(T(S))$$

where $T(S) = \{T1, Tt, Tt^2\} = \{t, 1 + t^2, 2t + t^3\}$. Since S is linearly independent and T is one-to-one, $T(S)$ is a basis for $R(T)$ by lemma 6.2.2 and hence $\text{rank}(T) = 3$. Since

$$\text{rank}(T) = 3 \neq 4 = \dim(P_3(\mathbb{R}))$$

we have that T is *not* onto by exercise 6.12.

Exercise 6.15

Consider the map $T : P_2(\mathbb{R}) \to R^2$ defined by

$$T\boldsymbol{f} = \begin{bmatrix} f(1) - f(2) \\ f(0) \end{bmatrix}$$

(1) Prove that T is linear.
(2) Find a basis for $N(T)$. What is $\text{nullity}(T)$?
(3) Is T one-to-one? Explain.
(4) Find a basis for $R(T)$. What is $\text{rank}(T)$?
(5) Is T onto? Explain.

Exercise 6.16

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that V is finite-dimensional.

(1) Prove that if $\text{nullity}(T) = 0$, then $\text{rank}(T) = \dim(V)$.
(2) Prove that if $\text{nullity}(T) = \dim(V)$, then $\text{rank}(T) = 0$.

6.3. Composition and Inverses

Definition 6.9

Let V, W, and Z be vector spaces and let $T : V \to W$ and $U : W \to Z$ be linear maps. We define the map $UT : V \to Z$ for $\boldsymbol{x} \in V$ by

$$(UT)\boldsymbol{x} = U(T\boldsymbol{x})$$

The map UT is called the *composition* of U and T.

Let V, W, and Z be vector spaces and let $T : V \to W$ and $U : W \to Z$ be linear maps. Using the fact that T and U are linear, we note that for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $a, b \in \mathbb{R}$ we have that

$$\begin{aligned}(UT)(a\boldsymbol{x} + b\boldsymbol{y}) &= U(T(a\boldsymbol{x} + b\boldsymbol{y})) \\ &= U(aT\boldsymbol{x} + bT\boldsymbol{y}) \\ &= aU(T\boldsymbol{x}) + bU(T\boldsymbol{y}) \\ &= a(UT)\boldsymbol{x} + b(UT)\boldsymbol{y}\end{aligned}$$

Thus, the map UT is linear. We have justified the following result.

Lemma 6.3.1

Let V, W, and Z be vector spaces and let $T : V \to W$ and $U : W \to Z$ be linear maps. Then the composition UT is linear.

Exercise 6.17

Let $T : V \to W$ and $U : W \to Z$ be linear maps. Prove that if UT is one-to-one, then T is one-to-one.

Let V, W, Y, and Z be vector spaces and let $S : V \to W$, $T : W \to Y$, and $U : Y \to Z$ be linear maps. For all $\boldsymbol{x} \in V$ we have that

$$(U(TS))\boldsymbol{x} = U((TS)\boldsymbol{x}) = U(T(S\boldsymbol{x})) = (UT)(S(\boldsymbol{x})) = ((UT)S)\boldsymbol{x}$$

We have shown that linear map composition is *associative*.

$$\boxed{U(TS) = (UT)S} \tag{6.7}$$

Let V be a vector space. The identity on V is the map $I_V : V \to V$ defined by $I_V\boldsymbol{x} = \boldsymbol{x}$. Note that I_V is a linear map.

Definition 6.10

Suppose V and W are vector spaces and let $T : V \to W$ be linear. A map $U : W \to V$ is called an *inverse* of T if U is linear and

$$UT = I_V, \qquad TU = I_W$$

If T has an inverse, then T is said to be *invertible*.

Suppose V and W are vector spaces and let $T : V \to W$ be linear. Suppose $U_1 : W \to V$ and $U_2 : W \to V$ are linear maps such that

$$U_1T = I_V, \qquad TU_1 = I_W, \qquad U_2T = I_V, \qquad TU_2 = I_W$$

Using equation (6.7) we note that

$$U_1 = U_1I_W = U_1(TU_2) = (U_1T)U_2 = I_VU_2 = U_2$$

and thus $U_1 = U_2$. We have shown the following result.

Lemma 6.3.2

Suppose V and W and vector spaces and let $T : V \to W$ be linear. If T is invertible, then its inverse is *unique*.

If T is invertible, we denote the unique inverse by T^{-1}.

Exercise 6.18

Suppose V and W are vector spaces and let $T : V \to W$ be linear. Prove that if T is invertible, then T^{-1} is invertible and $(T^{-1})^{-1} = T$.

Exercise 6.19

Let V, W, and Y be vector spaces and let $S : V \to W$ and $T : W \to Y$ be linear maps. Prove that if S and T are invertible, then TS is invertible and furthermore we have

$$(TS)^{-1} = S^{-1}T^{-1}$$

Suppose V and W are vector spaces and let $T : V \to W$ be a linear map that is invertible with inverse $T^{-1} : W \to V$ satisfying

$$T^{-1}T = I_V, \qquad TT^{-1} = I_W$$

Suppose $T\boldsymbol{v}_1 = T\boldsymbol{v}_2$ for some vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. Then we have that

$$\boldsymbol{v}_1 = I_V\boldsymbol{v}_1 = (T^{-1}T)\boldsymbol{v}_1 = T^{-1}(T\boldsymbol{v}_1) = T^{-1}(T\boldsymbol{v}_2) = (T^{-1}T)\boldsymbol{v}_2 = I_V\boldsymbol{v}_2 = \boldsymbol{v}_2$$

and thus T is one-to-one. Next let $\boldsymbol{w} \in W$. Then

$$\boldsymbol{w} = I_W\boldsymbol{w} = (TT^{-1})\boldsymbol{w} = T(T^{-1}\boldsymbol{w}) = T\boldsymbol{v}$$

where $\boldsymbol{v} = T^{-1}\boldsymbol{w} \in V$ and thus T is onto. We have justified the following result.

Lemma 6.3.3

Suppose V and W are vector spaces and let $T : V \to W$ be a linear map. If T is invertible, then T is one-to-one and onto.

Suppose V and W are vector spaces and let $T : V \to W$ be a linear map that is one-to-one and onto. We define an inverse $U : W \to V$ as follows. Let $\boldsymbol{w} = W$. Then since T is onto and one-to-one, there exists a *unique* $\boldsymbol{v} \in V$ such that $T\boldsymbol{v} = \boldsymbol{w}$. Define $U\boldsymbol{w} = \boldsymbol{v}$. Note that for all $\boldsymbol{v} \in V$ we have

$$(UT)\boldsymbol{v} = U(T\boldsymbol{v}) = U\boldsymbol{w} = \boldsymbol{v}$$

and for all $\boldsymbol{w} \in W$ we have

$$(TU)\boldsymbol{w} = T(U\boldsymbol{w}) = T\boldsymbol{v} = \boldsymbol{w}$$

We have shown that

$$UT = I_V, \qquad TU = I_W$$

It remains to show that U is linear. Let $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$ and let $a, b \in \mathbb{R}$. Let $\boldsymbol{v}_1 = U\boldsymbol{w}_1$ and $\boldsymbol{v}_2 = U\boldsymbol{w}_2$. Then $\boldsymbol{w}_1 = T\boldsymbol{v}_1$ and $\boldsymbol{w}_2 = T\boldsymbol{v}_2$. Since T is linear we have

$$\begin{aligned} U(a\boldsymbol{w}_1 + b\boldsymbol{w}_2) &= U(aT\boldsymbol{v}_1 + bT\boldsymbol{v}_2) \\ &= U(T(a\boldsymbol{v}_1 + b\boldsymbol{v}_2)) \\ &= (UT)(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \\ &= I_V(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \\ &= a\boldsymbol{v}_1 + b\boldsymbol{v}_2 \\ &= aU\boldsymbol{w}_1 + bU\boldsymbol{w}_2 \end{aligned}$$

When combined with lemma 6.3.3 we have justified the following theorem.

Theorem 6.1

Suppose V and W are vector spaces and let $T : V \to W$ be a linear map. Then T is invertible if and only if T is one-to-one and onto.

The next result follows from theorem 6.1 and exercise 6.14.

Lemma 6.3.4

Let V and W be finite-dimensional vector spaces and $T : V \to W$ be a linear map. If T is invertible, then $\dim(V) = \dim(W)$.

Exercise 6.20

Suppose V and W are vector spaces and $T : V \to W$ is a linear map that is invertible. Prove that V is finite-dimensional if and only if W is finite-dimensional.

6.4. The Rank-Nullity Theorem

Consider the map $T : P_2(\mathbb{R}) \to R^2$ defined by

$$T\boldsymbol{f} = \begin{bmatrix} f(1) - f(2) \\ f(0) \end{bmatrix}$$

In exercise 6.15 you showed that T is linear, $\text{nullity}(T) = 1$, and $\text{rank}(T) = 2$. The fact that

$$\dim(P_2(\mathbb{R})) = 3 = 1 + 2 = \text{nullity}(T) + \text{rank}(T)$$

is not an accident.

Let V and W be vector spaces and let $T : V \to W$ be a linear map and suppose that V is finite-dimensional. We will show that

$$\dim(V) = \text{nullity}(T) + \text{rank}(T)$$

Since $N(T)$ is a subspace of V we have that

$$0 \leq \text{nullity}(T) \leq \dim(V)$$

Exercise 6.16 covers the special cases $\text{nullity}(T) = 0$ and $\text{nullity}(T) = \dim(V)$ so suppose $\text{nullity}(T) = k$ where $0 < k < \dim(V)$. Let $L = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be a basis for $N(T)$. Since L is a linearly independent subset of V it can be extended to a basis for V by theorem 5.6. Let $n = \dim(V)$ and let $\beta = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n\}$ be an extended basis for V. We will show that $S = \{T\boldsymbol{u}_{k+1}, \ldots, T\boldsymbol{u}_n\}$ contains $n - k$ vectors and is a basis for $R(T)$. This will show that $\text{rank}(T) = \dim(V) - k$ and we will have that

$$\dim(V) = k + (\dim(V) - k) = \text{nullity}(T) + \text{rank}(T)$$

First suppose that

$$a_{k+1}T\boldsymbol{u}_{k+1} + \cdots + a_nT\boldsymbol{u}_n = \boldsymbol{0}$$

Since T is linear we have

$$T(a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n) = \boldsymbol{0}$$

Thus

$$a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n \in N(T)$$

Since L is a basis for $N(T)$ we can write

$$a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n = a_1\boldsymbol{u}_1 + \cdots a_k\boldsymbol{u}_k$$

After rearranging we get

$$a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n - a_1\boldsymbol{u}_1 - \cdots - a_k\boldsymbol{u}_k = \boldsymbol{0}$$

But since β is a basis for V we must have that

$$a_1 = \cdots = a_k = a_{k+1} = \cdots = a_n = 0$$

Thus, S is linear independent set of $n-k$ vectors. To check to see if S generates $R(T)$ we apply lemma 6.2.4 and calculate

$$\begin{aligned} R(T) &= \operatorname{span}(T(\beta)) \\ &= \operatorname{span}(\{T\boldsymbol{u}_1, \ldots, T\boldsymbol{u}_k, T\boldsymbol{u}_{k+1}, \ldots T\boldsymbol{u}_n\}) \\ &= \operatorname{span}(\{\mathbf{0}, \ldots, \mathbf{0}, T\boldsymbol{u}_{k+1}, \ldots T\boldsymbol{u}_n\}) \\ &= \operatorname{span}(\{T\boldsymbol{u}_{k+1}, \ldots T\boldsymbol{u}_n\}) = \operatorname{span}(S) \end{aligned}$$

We have justified the following theorem.

Theorem 6.2: Rank-Nullity

Let V and W be vector spaces, and let $T : V \to W$ be linear. If V is finite-dimensional then

$$\dim(V) = \operatorname{rank}(T) + \operatorname{nullity}(T)$$

The next corollary follows from theorem 6.2. We leave the proof as an exercise.

Corollary 6.4.1

Let $A \in \mathbb{R}^{m\times n}$. Then

$$n = \operatorname{rank}(A) + \operatorname{nullity}(A)$$

Exercise 6.21

Prove corollary 6.4.1.

Consider the set $W = \{\boldsymbol{x} \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0\}$. Since $W = N(A)$ where $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$, W is a subspace of $\mathbb{R}^4$. By inspection, $\operatorname{rank}(A) = 1$ and thus $\operatorname{nullity}(A) = \dim(W) = 4 - 1 = 3$ by corollary 6.4.1. We note that

$$\beta = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a linearly independent set of $\dim(W)$ vectors in W and hence is a basis for W by corollary 5.3.2.

Exercise 6.22

Consider the set $W = \{\boldsymbol{x} \in \mathbb{R}^4 : x_1 - x_2 + x_3 - x_4 = 0\}$.

(1) Show that W is a subspace of $\mathbb{R}^4$.
(2) Calculate $\dim(W)$.
(3) Find a basis for W.

Exercise 6.23

Consider the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

(1) Calculate $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.
(2) Find a basis for $N(A)$.

Exercise 6.24

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 0 & -1 & 1 \end{bmatrix}$.

(1) Calculate $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.
(2) Find a basis for $N(A)$.

Exercise 6.25

Consider the matrix $A = \begin{bmatrix} 1 & 3 & 2 \\ 1 & 0 & 2 \\ 0 & 3 & 0 \end{bmatrix}$.

(1) Calculate $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.
(2) Find a basis for $N(A)$.

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. Then we have that

$$N(A) = N(C)$$

by lemma 3.4.4. Thus $\operatorname{nullity}(A) = \operatorname{nullity}(C)$ and

$$\operatorname{rank}(A) = n - \operatorname{nullity}(A) = n - \operatorname{nullity}(C) = \operatorname{rank}(C)$$

by corollary 6.4.1. We have justified the following result.

Corollary 6.4.2

Let $A \in \mathbb{R}^{m\times n}$ and let $C \in \mathbb{R}^{m\times n}$ be the matrix resulting from applying any number of elementary row operations to A. Then we have that

$$\operatorname{rank}(A) = \operatorname{rank}(C)$$

Consider the matrix $A = \begin{bmatrix} 1 & 4 & 3 \\ -4 & -7 & -3 \\ 3 & 3 & 0 \end{bmatrix}$. Adding 4 times the first row to the second row and -3 times the first row to the third row gives the matrix $C = \begin{bmatrix} 1 & 4 & 3 \\ 0 & 9 & 9 \\ 0 & -9 & -9 \end{bmatrix}$. By corollary 6.4.2 we have that $\operatorname{rank}(A) = \operatorname{rank}(C) = 2$. Thus by corollary 6.4.1 we have $\operatorname{nullity}(A) = 3 - \operatorname{rank}(A) = 3 - 2 = 1$. By inspection we have $\beta = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$ is a linearly independent set of $\operatorname{nullity}(A)$ vectors in $N(A)$ and hence β is a basis for $N(A)$ by corollary 5.3.2.

Exercise 6.26

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 1 & 1 & 1 \\ -1 & 4 & 7 & 1 \end{bmatrix}$.

(1) Calculate $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.
(2) Find a basis for $N(A)$.

Exercise 6.27

Let $A \in \mathbb{R}^{m\times n}$. If $\operatorname{rank}(A) = m$ and $\operatorname{nullity}(A) = 0$ what can you say about m and n? Explain.

Exercise 6.28

Let $A \in \mathbb{R}^{15\times 20}$ and suppose that $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for *all* possible right hand sides $\boldsymbol{b}$. Calculate $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.

Exercise 6.29

Let V and W be vector spaces, and let $T : V \to W$ be linear. Suppose V is finite-dimensional. Prove that T is one-to-one if and only if $\operatorname{rank}(T) = \dim(V)$.

Exercise 6.30

Let $T : V \to W$ be a linear map and suppose that V and W are finite-dimensional vector spaces such that $\dim(V) = \dim(W)$. Prove that the following statements are equivalent.

(1) T is one-to-one.
(2) T is onto.
(3) T is invertible.

Exercise 6.31

Let V and W be finite-dimensional vector spaces. Let $T : V \to W$ and $U : W \to V$ be linear and assume that $\dim(V) = \dim(W)$. Prove that if $UT = I_V$, then $TU = I_W$.

Let $A \in \mathbb{R}^{m\times n}$ and $P \in \mathbb{R}^{p\times m}$ such that $N(P) = \{\mathbf{0}\}$. Suppose $\boldsymbol{x} \in N(PA)$. Then $(PA)\boldsymbol{x} = \mathbf{0}$. Hence we have

$$\mathbf{0} = (PA)\boldsymbol{x} = P(A\boldsymbol{x})$$

and $A\boldsymbol{x} \in N(P) = \{\mathbf{0}\}$. Thus $A\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{x} \in N(A)$. Next suppose $\boldsymbol{x} \in N(A)$. Then $A\boldsymbol{x} = \mathbf{0}$. Hence we have

$$(PA)\boldsymbol{x} = P(A\boldsymbol{x}) = P\mathbf{0} = \mathbf{0}$$

Since $N(A) = N(PA)$ we have that $\operatorname{nullity}(A) = \operatorname{nullity}(PA)$. By corollary 6.4.1 we have that

$$\operatorname{rank}(A) = n - \operatorname{nullity}(A) = n - \operatorname{nullity}(PA) = \operatorname{rank}(PA)$$

We have justified the following lemma.

Lemma 6.4.1

Let $A \in \mathbb{R}^{m\times n}$ and $P \in \mathbb{R}^{p\times m}$ such that $N(P) = \{\mathbf{0}\}$. Then

$$\operatorname{rank}(A) = \operatorname{rank}(PA)$$

Exercise 6.32

Let $A \in \mathbb{R}^{m\times n}$ and $Q \in \mathbb{R}^{n\times q}$ such that $\operatorname{rank}(Q) = n$. Prove that $\operatorname{col}(A) = \operatorname{col}(AQ)$ and hence $\operatorname{rank}(A) = \operatorname{rank}(AQ)$.

The following theorem follows from lemma 6.4.1 and exercise 6.32.

Theorem 6.3

Let $A \in \mathbb{R}^{m\times n}$, $P \in \mathbb{R}^{p\times m}$ such that $N(P) = \{\mathbf{0}\}$, and $Q \in \mathbb{R}^{n\times q}$ such that $\operatorname{rank}(Q) = n$. Then $\operatorname{rank}(A) = \operatorname{rank}(PAQ)$.

Exercise 6.33

Let $A \in \mathbb{R}^{m\times n}$. Suppose that $Q \in \mathbb{R}^{n\times n}$ is invertible and that $P \in \mathbb{R}^{m\times m}$ is invertible. Prove that $\operatorname{rank}(A) = \operatorname{rank}(PAQ)$.

Exercise 6.34

Let V, W, Y, and Z be finite-dimensional vector spaces and let $S : V \to W$, $T : W \to Y$, and $U : Y \to Z$ be linear maps. Prove that

(1) if S is onto, then $\operatorname{rank}(T) = \operatorname{rank}(TS)$.
(2) if U is one-to-one, then $\operatorname{rank}(T) = \operatorname{rank}(UT)$.
(3) if S is onto and U is one-to-one, then $\operatorname{rank}(T) = \operatorname{rank}(UTS)$.

Exercise 6.35

Let $\boldsymbol{x} \in \mathbb{R}^n$. Prove that $\boldsymbol{x}^T\boldsymbol{x} = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$.

Exercise 6.36

Let $A \in \mathbb{R}^{m\times n}$. Prove that if $A^TA\boldsymbol{x} = \mathbf{0}$, then $A\boldsymbol{x} = \mathbf{0}$. *Hint:* Calculate $(A\boldsymbol{x})^TA\boldsymbol{x}$.

Exercise 6.37

Let $A \in \mathbb{R}^{m\times n}$.

(1) Prove that $N(A) = N(A^TA)$.
(2) Prove that $\operatorname{rank}(A) = \operatorname{rank}(A^TA)$.

Exercise 6.38

Let $A \in \mathbb{R}^{m\times n}$.

(1) Prove that $\operatorname{rank}(A) = \operatorname{rank}(AA^T)$.
(2) Prove that $\operatorname{col}(AA^T) \subseteq \operatorname{col}(A)$.
(3) Conclude that $\operatorname{col}(AA^T) = \operatorname{col}(A)$ and hence $\operatorname{col}(A^TA) = \operatorname{col}(A^T)$.
(4) Conclude that for any $\boldsymbol{b} \in \mathbb{R}^m$ the linear system $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ is consistent.

Exercise 6.39

Let $A \in \mathbb{R}^{m\times n}$.

(1) Prove that A^TA is invertible if and only if $\operatorname{rank}(A) = n$.
(2) Prove that AA^T is invertible if and only if $\operatorname{rank}(A) = m$.

CHAPTER 7

Matrix Representations and Change of Coordinates

7.1. Coordinate Maps

Let V be a finite-dimensional vector space. An *ordered basis* is a basis $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ where the sequence of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ determines the ordering. Suppose $V = \mathbb{R}^3$. Let $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ and $\gamma = \{\boldsymbol{e}_2, \boldsymbol{e}_1, \boldsymbol{e}_3\}$. Then β and γ are the same basis for V but β and γ are *not* the same as ordered bases. We call $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ the standard ordered basis for $\mathbb{R}^n$ and $\gamma = \{1, t, \ldots, t^n\}$ the standard ordered basis for $P_n(\mathbb{R})$.

Let V be a finite-dimensional vector space and let $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be an ordered basis of n vectors in V. We define the map $T : \mathbb{R}^n \to V$ by

$$T\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n$$

By exercise 6.4, the map T is linear. Suppose $T\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x} \in \mathbb{R}^n$. Then we have

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n = \boldsymbol{0}$$

Since β is a basis, we must have $x_1 = x_2 = \cdots = x_n = 0$ and thus $N(T) = \{\boldsymbol{0}\}$ and T is one-to-one. Since $\dim(\mathbb{R}^n) = \dim(V)$ we have that T is onto and invertible by exercise 6.30.

Definition 7.1

Let V be a finite-dimensional vector space and let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis of n vectors in V. Let $T : \mathbb{R}^n \to V$ be the linear map defined by

$$T\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n$$

Then T is invertible and we define the *coordinate map relative to* β by

$$\phi_\beta = T^{-1}$$

Let V be a finite-dimensional vector space and let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis of n vectors in V. Note that $\phi_\beta : V \to \mathbb{R}^n$ is linear since it is the inverse of a linear map. Furthermore, ϕ_β is invertible, one-to-one, and onto by exercise 6.18 and theorem 6.1.

For all i in $1, \ldots, n$ we have

$$\phi_\beta\boldsymbol{v}_i = \boldsymbol{e}_i, \qquad \phi_\beta^{-1}\boldsymbol{e}_i = \boldsymbol{v}_i \tag{7.1}$$

We also use the notation

$$[\boldsymbol{v}]_\beta = \phi_\beta\boldsymbol{v}$$

for the coordinates of a specific vector $\boldsymbol{v}$ relative to β.

Exercise 7.1

Let V and W be finite dimensional vector spaces. We say that V is *isomorphic* to W if there exists a linear map $T : V \to W$ that is invertible. Such a linear map is called an *isomorphism* from V onto W.

(1) Prove that if $\dim(V) = n$, then V is isomorphic to $\mathbb{R}^n$.
(2) Prove that V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

Let $\beta = \{1 + 2t + t^2, 3 + t^2, t + t^2\}$. We leave it as an exercise to show that β is an ordered basis for $P_2(\mathbb{R})$. Suppose $\boldsymbol{f} \in P_2(\mathbb{R})$ and that $[\boldsymbol{f}]_\beta = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Applying ϕ_β^{-1} to both sides and using equation (7.1) we note that

$$\boldsymbol{f} = 1(1 + 2t + t^2) + (-1)(3 + t^2) + 1(t + t^2) = t^2 + 3t - 2$$

To calculate $[2t^2 + 3t + 5]_\beta$, we find c_1, c_2, c_3 such that

$$2t^2 + 3t + 5 = c_1(1 + 2t + t^2) + c_2(3 + t^2) + c_3(t + t^2)$$

by solving the system of equations

$$\begin{aligned} c_1 + c_2 + c_3 &= 2 \\ 2c_1 + c_3 &= 3 \\ c_1 + 3c_2 &= 5 \end{aligned}$$

The unique solution is

$$c_1 = 2, \qquad c_2 = 1, \qquad c_3 = -1$$

and hence

$$[2t^2 + 3t + 5]_\beta = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

Exercise 7.2

Show that

$$\beta = \{t^2, 1, 2t\}, \qquad \gamma = \{t^2 - t + 1, t + 1, t^2 + 1\}$$

are ordered bases for $P_2(\mathbb{R})$. Find $\boldsymbol{f}, \boldsymbol{g} \in P_2(\mathbb{R})$ given that $[\boldsymbol{f}]_\beta = [\boldsymbol{g}]_\gamma = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$. Finally, calculate $[t^2 + 2t + 2]_\beta$ and $[t^2 + 2t + 2]_\gamma$.

Exercise 7.3

Let V be a vector space such that $\dim(V) = n$ and let β be an ordered basis for V.

(1) Prove that γ is a basis for V if and only if $\phi_\beta(\gamma)$ is a basis for $\mathbb{R}^n$.

(2) Prove that α is a basis for $\mathbb{R}^n$ if and only if $\phi_\beta^{-1}(\alpha)$ is a basis for V.

7.2. Matrix Representations

Let V and W be finite-dimensional vector spaces and $T : V \to W$ be a linear map. Suppose $\dim(V) = n$ and $\dim(W) = m$ and let $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be an ordered basis for V and $\gamma = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}$ be an ordered basis for W. Let $\phi_\beta : V \to \mathbb{R}^n$ and $\phi_\gamma : W \to \mathbb{R}^m$ be the associated coordinate maps. We will study the linear map T by considering how it maps coordinates in $\mathbb{R}^n$ to coordinates in $\mathbb{R}^m$. In particular, consider the linear map $\phi_\gamma T \phi_\beta^{-1} : \mathbb{R}^n \to \mathbb{R}^m$. Since $\phi_\gamma T \phi_\beta^{-1}$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$\phi_\gamma T \phi_\beta^{-1} \boldsymbol{x} = A\boldsymbol{x} \tag{7.2}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ by lemma 6.1.3. To find the matrix A we let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary and calculate

$$\begin{aligned}
\phi_\gamma T \phi_\beta^{-1}\boldsymbol{x} &= \phi_\gamma T \phi_\beta^{-1}(x_1\boldsymbol{e}_1 + \cdots x_n\boldsymbol{e}_n) \\
&= x_1\phi_\gamma T\phi_\beta^{-1}\boldsymbol{e}_1 + \cdots + x_n\phi_\gamma T\phi_\beta^{-1}\boldsymbol{e}_n \\
&= x_1\phi_\gamma T\boldsymbol{v}_1 + \cdots + x_n\phi_\gamma T\boldsymbol{v}_n \\
&= x_1[T\boldsymbol{v}_1]_\gamma + \cdots + x_n[T\boldsymbol{v}_n]_\gamma \\
&= \begin{bmatrix}[T\boldsymbol{v}_1]_\gamma & \cdots & [T\boldsymbol{v}_n]_\gamma\end{bmatrix}\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix} \\
&= A\boldsymbol{x} \qquad (7.3)
\end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ is the matrix

$$A = \begin{bmatrix}[T\boldsymbol{v}_1]_\gamma & \cdots & [T\boldsymbol{v}_n]_\gamma\end{bmatrix} \qquad (7.4)$$

Definition 7.2

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces V and W where $\dim(V) = n$ and $\dim(W) = m$. Let $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be an ordered basis for V and and γ be an ordered basis for W. We define the *matrix representation of T relative to β and γ* by

$$[T]_\beta^\gamma = \begin{bmatrix}[T\boldsymbol{v}_1]_\gamma & [T\boldsymbol{v}_2]_\gamma & \cdots & [T\boldsymbol{v}_n]_\gamma\end{bmatrix} \qquad (7.5)$$

By equations (7.3) and (7.5) we have that

$$[T]_\beta^\gamma\boldsymbol{x} = \phi_\gamma T\phi_\beta^{-1}\boldsymbol{x} \qquad (7.6)$$

for all $\boldsymbol{x}$ in $\mathbb{R}^n$. Let $\boldsymbol{v} \in V$. Using equation (7.6) with $\boldsymbol{x} = [\boldsymbol{v}]_\beta$ gives

$$[T]_\beta^\gamma[\boldsymbol{v}]_\beta = \phi_\gamma T\phi_\beta^{-1}[\boldsymbol{v}]_\beta = [T\boldsymbol{v}]_\gamma$$

We summarize this result in the next lemma.

Lemma 7.2.1

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces, and let β and γ be ordered bases for V and W. For all $\boldsymbol{v} \in V$ we have

$$[T\boldsymbol{v}]_\gamma = [T]_\beta^\gamma[\boldsymbol{v}]_\beta \qquad (7.7)$$

Let $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ be the standard ordered basis for $\mathbb{R}^2$ and let $\gamma = \{\boldsymbol{e}_3, \boldsymbol{e}_2, \boldsymbol{e}_1\}$ be an ordered basis for $\mathbb{R}^3$. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map defined by $T\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}x_1+3x_2\\0\\2x_1-4x_2\end{bmatrix}$. We calculate

$$\begin{aligned}
[T]_\beta^\gamma &= \begin{bmatrix}\left[T\begin{bmatrix}1\\0\end{bmatrix}\right]_\gamma & \left[T\begin{bmatrix}0\\1\end{bmatrix}\right]_\gamma\end{bmatrix} \\
&= \begin{bmatrix}\left[\begin{bmatrix}1\\0\\2\end{bmatrix}\right]_\gamma & \left[\begin{bmatrix}3\\0\\-4\end{bmatrix}\right]_\gamma\end{bmatrix} \\
&= \begin{bmatrix}2 & -4\\0 & 0\\1 & 3\end{bmatrix}
\end{aligned}$$

Let $\boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ in $\mathbb{R}^2$. We verify equation (7.7) by noting that

$$[T\boldsymbol{x}]_\gamma = \left[\begin{bmatrix} -2 \\ 0 \\ 6 \end{bmatrix}\right]_\gamma = \begin{bmatrix} 6 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ 0 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = [T]_\beta^\gamma [\boldsymbol{x}]_\beta$$

Exercise 7.4

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ x_1 \\ 2x_1 + x_2 \end{bmatrix}$. Let $\beta = \{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ be the standard ordered basis for $\mathbb{R}^2$. Verify that $\gamma = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \right\}$ is an ordered basis for $\mathbb{R}^3$. Calculate $[T]_\beta^\gamma$. Verify that $[T\boldsymbol{x}]_\gamma = [T]_\beta^\gamma [\boldsymbol{x}]_\beta$ for $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Exercise 7.5

Let $T : \mathbb{R}^2 \to P_2(\mathbb{R})$ be the map defined by

$$T \begin{bmatrix} a \\ b \end{bmatrix} = (a+b)(t+1)^2 + a(t+1) + b$$

Show that T is linear. Let $\beta = \left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$ and $\gamma = \{2t^2, t, -1\}$ be ordered bases for $\mathbb{R}^2$ and $P_2(\mathbb{R})$ respectively. Calculate $[T]_\beta^\gamma$.

Exercise 7.6

Let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. Verify that

$$\gamma = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is an ordered basis for $\mathbb{R}^{2\times 2}$. Verify that $T : P_2(\mathbb{R}) \to \mathbb{R}^{2\times 2}$ defined by

$$T\boldsymbol{f} = \begin{bmatrix} f'(0) & 2f(1) \\ 0 & f''(3) \end{bmatrix}$$

is linear. Calculate $[T]_\beta^\gamma$.

Exercise 7.7

Let V and W be finite dimensional vector spaces with ordered bases β and γ, respectively. Suppose $T : V \to W$ and $U : V \to W$ are linear maps. Prove that $[T]_\beta^\gamma = [U]_\beta^\gamma$ if and only if $T = U$.

Exercise 7.8

Let $T : V \to W$ and $U : V \to W$ be linear maps between finite-dimensional vector spaces V and W. Let β be an ordered basis for V and γ be an ordered basis for W. Let $a \in \mathbb{R}$. The map $T + aU$ is defined for $\boldsymbol{x} \in V$ by $(T + aU)\boldsymbol{x} = T\boldsymbol{x} + aU\boldsymbol{x}$. By exercise 6.3 we have that $T + aU$ is a linear map. Prove that

$$[T + aU]_\beta^\gamma = [T]_\beta^\gamma + a[U]_\beta^\gamma$$

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces V and W, and let β and γ be ordered bases for V and W, respectively. Recall that by equation (7.6) we have that

$$[T]_\beta^\gamma \boldsymbol{x} = \phi_\gamma T \phi_\beta^{-1} \boldsymbol{x}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. By exercise 6.34 and lemma 6.2.5 we therefore have that

$$\operatorname{rank}(T) = \operatorname{rank}(\phi_\gamma T \phi_\beta^{-1}) = \operatorname{rank}([T]_\beta^\gamma)$$

We summarize this result in the following theorem.

Theorem 7.1

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces V and W, and let β and γ be ordered bases for V and W, respectively. Then

$$\operatorname{rank}(T) = \operatorname{rank}([T]_\beta^\gamma)$$

Exercise 7.9

Let V and W be finite-dimensional vector spaces and $T : V \to W$ be linear. Prove that if β and β' are ordered bases for V and γ and γ' are ordered bases for W, then

$$\operatorname{rank}([T]_\beta^\gamma) = \operatorname{rank}([T]_{\beta'}^{\gamma'})$$

Let V and W be finite-dimensional vector spaces with ordered bases β and γ respectively. Let $T : V \to W$ be linear and suppose that

$$[T]_\beta^\gamma = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & -1 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

To calculate rank of $[T]_\beta^\gamma$ we first do row operations on $[T]_\beta^\gamma$ which preserve the rank. Adding -1 times row 1 to row 2 and -1 times row 1 to row 3 gives

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

Adding 2 times row 3 to row 2 and -1 times row 3 to row 1 gives

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

Adding row 2 to row 1 and multiply row 2 by -1 gives

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

Thus we have $\operatorname{rank}([T]_\beta^\gamma) = \operatorname{rank}(C) = 3$. Since $\operatorname{rank}(T) = \operatorname{rank}([T]_\beta^\gamma)$ by theorem 7.1, we have that $\operatorname{rank}(T) = 3$. Since $\dim(V) = 4$, we have that

$$\operatorname{nullity}(T) = 4 - \operatorname{rank}(T) = 4 - 3 = 1$$

by the rank-nullity theorem. Since $\operatorname{nullity}(T) > 0$, T is not one-to-one. Since $\operatorname{rank}(T) = 3 = \dim(W)$, we have that T is onto.

Exercise 7.10

Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Let $T : V \to W$ be linear and suppose that

$$[T]_\beta^\gamma = \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 4 \\ 2 & 0 & 2 \\ 2 & 4 & -5 \end{bmatrix}$$

Calculate rank(T) and nullity(T). Is T one-to-one? Is T onto? Explain.

Definition 7.3

Let V be a finite-dimensional vector space with ordered basis β and suppose $T : V \to V$ is a linear map. We define

$$[T]_\beta = [T]_\beta^\beta$$

Exercise 7.11

Let V be a vector space such that $\dim(V) = n$. Let β be an ordered basis for V. Let $I_V : V \to V$ be the identity map on V. Prove that $[I_V]_\beta = I_n$

Exercise 7.12

Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $(T\boldsymbol{f})(t) = f(t) + f'(t) + f''(t)$. Show that T is linear. Let $\beta = \{1, t, t^2\}$ be the standard ordered basis of $P_2(\mathbb{R})$. Calculate $[T]_\beta$. Calculate rank(T) and nullity(T). Is T one-to-one? Is T onto? Is T invertible? Explain.

Let V, W, and Z be finite-dimensional vector spaces with ordered bases β, γ, and α, respectively. Let $T : V \to W$ and $U : W \to Z$ be linear maps. Suppose $\dim(V) = n$ and let $\beta = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be an ordered basis for V. We calculate

$$\begin{aligned} [UT]_\beta^\alpha &= \begin{bmatrix} [(UT)\boldsymbol{v}_1]_\alpha & \cdots & [(UT)\boldsymbol{v}_n]_\alpha \end{bmatrix} \\ &= \begin{bmatrix} [U(T\boldsymbol{v}_1)]_\alpha & \cdots & [U(T\boldsymbol{v}_n)]_\alpha \end{bmatrix} \\ &= \begin{bmatrix} [U]_\gamma^\alpha [T\boldsymbol{v}_1]_\gamma & \cdots & [U]_\gamma^\alpha [T\boldsymbol{v}_n]_\gamma \end{bmatrix} \\ &= [U]_\gamma^\alpha \begin{bmatrix} [T\boldsymbol{v}_1]_\gamma & \cdots & [T\boldsymbol{v}_n]_\gamma \end{bmatrix} \\ &= [U]_\gamma^\alpha [T]_\beta^\gamma \end{aligned}$$

where we used lemma 7.2.1 to justify that $[U(T\boldsymbol{v}_i)]_\alpha = [U]_\gamma^\alpha [T\boldsymbol{v}_i]_\gamma$. We summarize this result in the following theorem.

Theorem 7.2

Let V, W, and Z be finite-dimensional vector spaces with ordered bases β, γ, and α, respectively. Let $T : V \to W$ and $U : W \to Z$ be linear maps. Then

$$[UT]_\beta^\alpha = [U]_\gamma^\alpha [T]_\beta^\gamma \tag{7.8}$$

Exercise 7.13

Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ and $U : P_2(\mathbb{R}) \to \mathbb{R}^3$ be the maps defined by

$$(T\boldsymbol{f})(t) = f'(t)(3+t) + 2f(t), \qquad U(a + bt + ct^2) = \begin{bmatrix} a+b \\ c \\ a-b \end{bmatrix}$$

Prove that T and U are linear maps. Let $\beta = \{1, t, t^2\}$ and $\gamma = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the standard ordered bases of $P_2(\mathbb{R})$ and $\mathbb{R}^3$, respectively. Compute $[U]_\beta^\gamma$, $[T]_\beta$, and $[UT]_\beta^\gamma$ directly. Use equation 7.8 to calculate $[UT]_\beta^\gamma$ and verify theorem 7.2.

Let V and W be finite-dimensional vector spaces with ordered bases β and γ respectively and suppose that $\dim(V) = \dim(W) = n$. Let $T : V \to W$ be a linear map. Since $\operatorname{rank}(T) = \operatorname{rank}([T]_\beta^\gamma)$ we have that T is invertible if and only if $[T]_\beta^\gamma$ is invertible by exercises 5.19 and 6.30. If T is invertible, then by theorem 7.2 and exercise 7.11 we have that

$$[T^{-1}]_\gamma^\beta [T]_\beta^\gamma = [T^{-1}T]_\beta = [I_V]_\beta = I_n$$

$$[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = [TT^{-1}]_\gamma = [I_W]_\gamma = I_n$$

We summarize these results in the following theorem.

Theorem 7.3

Let V and W be finite-dimensional vector spaces with ordered bases β and γ and suppose $\dim(V) = \dim(W) = n$. Let $T : V \to W$ be a linear map. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible. Furthermore, if T is invertible, then

$$([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta \tag{7.9}$$

Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be the map defined by $(T\boldsymbol{f})(t) = f(t) - 2f'(t)$. We leave it as an exercise to verify that T is linear. Let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. We calculate

$$T1 = 1, \qquad Tt = t - 2, \qquad Tt^2 = t^2 - 4t$$

$$[T]_\beta = \begin{bmatrix} [T1]_\beta & [Tt]_\beta & [Tt^2]_\beta \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$$

Using standard matrix inverse techniques we calculate

$$[T^{-1}]_\beta = ([T]_\beta)^{-1} = \begin{bmatrix} 1 & 2 & 8 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Next we calculate

$$[T^{-1}(2-t^2)]_\beta = [T^{-1}]_\beta [2-t^2]_\beta = \begin{bmatrix} 1 & 2 & 8 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -6 \\ -4 \\ -1 \end{bmatrix}$$

Applying ϕ_β^{-1} to both sides gives

$$T^{-1}(2-t^2) = -t^2 - 4t - 6$$

Finally we note that

$$T(-t^2 - 4t - 6) = (-t^2 - 4t - 6) - 2(-2t - 4) = 2 - t^2$$

as expected.

Exercise 7.14

Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $(T\boldsymbol{f})(t) = f(t) + f'(t) + f''(t)$. In exercise 7.12 you showed that T is a linear map. Let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. Calculate $[T]_\beta$ and $[T^{-1}]_\beta$. Use $[T^{-1}]_\beta$ to calculate $T^{-1}(t^2 - 2t + 1)$. Verify your answer.

7.3. Change of Coordinates

Let V be a finite-dimensional vector space with ordered bases β and γ. Let $T : V \to V$ be a linear map. For any $\boldsymbol{v} \in V$ we have that

$$[\boldsymbol{v}]_\gamma = [I_V \boldsymbol{v}]_\gamma = [I_V]_\beta^\gamma [\boldsymbol{v}]_\beta \tag{7.10}$$

We say that the matrix $[I_V]_\beta^\gamma$ changes coordinates relative to β into coordinates relative to γ. Similarily, for any $\boldsymbol{v} \in V$ we have that

$$[\boldsymbol{v}]_\beta = [I_V \boldsymbol{v}]_\beta = [I_V]_\gamma^\beta [\boldsymbol{v}]_\gamma \tag{7.11}$$

and we say that the matrix $[I_V]_\gamma^\beta$ changes coordinates relative to γ into coordinates relative to β.

Let $\beta = \{t^2+t+4, 4t^2-3t+2, 2t^2+3\}$ and $\gamma = \{t^2-t+1, t+1, t^2+1\}$ be subsets of $P_2(\mathbb{R})$. We leave it as an exercise to show that β and γ are ordered bases for $P_2(\mathbb{R})$. To calculate $[I_V]_\beta^\gamma$, we first calculate the coordinates of the vectors in β relative to γ by solving systems of equations. We note that

$$[t^2+t+4]_\gamma = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}, \qquad [4t^2-3t+2]_\gamma = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \qquad [2t^2+3]_\gamma = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Next, we note that by definition

$$[I_V]_\beta^\gamma = \big[[t^2+t+4]_\gamma \quad [4t^2-3t+2]_\gamma \quad [2t^2+3]_\gamma\big] = \begin{bmatrix} 2 & 1 & 1 \\ 3 & -2 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

Exercise 7.15

Let $\beta = \left\{ \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$ and $\gamma = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$. Verify that β and γ are ordered bases for $\mathbb{R}^2$. Find a matrix $Q \in \mathbb{R}^{2\times 2}$ that changes coordinates relative to β into coordinates relative to γ.

By theorem 7.3 we have that

$$([I_V]_\beta^\gamma)^{-1} = [I_V^{-1}]_\gamma^\beta = [I_V]_\gamma^\beta \tag{7.12}$$

By theorem 7.2 we have that

$$[T]_\beta = [I_V\, T\, I_V]_\beta = [I_V\, T]_\gamma^\beta [I_V]_\beta^\gamma = [I_V]_\gamma^\beta [T]_\gamma^\gamma [I_V]_\beta^\gamma \tag{7.13}$$

and

$$[T]_\gamma = [I_V\, T\, I_V]_\gamma = [I_V\, T]_\beta^\gamma [I_V]_\gamma^\beta = [I_V]_\beta^\gamma [T]_\beta^\beta [I_V]_\gamma^\beta \tag{7.14}$$

Let V be a finite-dimensional vector space with ordered bases β and γ. Let $T : V \to V$ be a linear map. Suppose that for some $\boldsymbol{v} \in V$ we have

$$[\boldsymbol{v}]_\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad [T]_\beta = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}, \qquad [I_V]_\gamma^\beta = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}$$

Using equation (7.10) we calculate

$$[\boldsymbol{v}]_\gamma = [I_V]_\beta^\gamma [\boldsymbol{v}]_\beta = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

Using equation (7.12) we calculate

$$[I_V]_\beta^\gamma = ([I_V]_\gamma^\beta)^{-1} = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}^{-1} = \frac{1}{5}\begin{bmatrix} 1 & 1 \\ -3 & 2 \end{bmatrix}$$

Using equation (7.14) we calculate

$$[T]_\gamma = [I_V]_\beta^\gamma [T]_\beta^\beta [I_V]_\gamma^\beta = \frac{1}{5}\begin{bmatrix} 1 & 1 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 1 & 1 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 8 & 1 \\ 12 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$$

Exercise 7.16

Let V be a finite-dimensional vector space with ordered bases β and γ. Let $T : V \to V$ be a linear map. Given that $[T]_\beta = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$ and $[I_V]_\gamma^\beta = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}$, calculate $[T]_\gamma$.

CHAPTER 8

Diagonalizable Operators

8.1. Eigenvalues and Eigenvectors

Let V be a vector space. We call a linear map $T : V \to V$ a *linear operator* on V. A matrix $D \in \mathbb{R}^{n\times n}$

$$D = \begin{bmatrix} d_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{nn} \end{bmatrix}$$

is called *diagonal* if $d_{ij} = 0$ for $i \neq j$.

Definition 8.1

A linear operator $T : V \to V$ on a finite-dimensional vector space V is called *diagonalizable* if there is an ordered basis γ for V such that $[T]_\gamma$ is a diagonal matrix.

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by

$$(T\boldsymbol{f})(t) = f(t) - 2tf'(t) + (t^2+1)f''(t)$$

Let $\gamma = \{1, t, 1-t^2\}$. We calculate

$$[T]_\gamma = \big[[T1]_\gamma \quad [Tt]_\gamma \quad [T(1-t^2)]_\gamma\big] = \big[[1]_\gamma \quad [-t]_\gamma \quad [t^2-1]_\gamma\big] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = D$$

Thus, we have that T is a diagonalizable operator. Since

$$T1 = (1)1, \qquad Tt = (-1)t, \qquad T(1-t^2) = (-1)(1-t^2)$$

we have that for each $\boldsymbol{v} \in \gamma$ that $T\boldsymbol{v} = \lambda\boldsymbol{v}$ where $\lambda \in \mathbb{R}$.

Definition 8.2

Let $T : V \to V$ be a linear operator on a vector space V. A scalar $\lambda \in \mathbb{R}$ is called an *eigenvalue* of T if there exists a *nonzero* vector $\boldsymbol{v} \in V$, called an *eigenvector* of T, such that

$$T\boldsymbol{v} = \lambda\boldsymbol{v} \tag{8.1}$$

Exercise 8.1

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map defined by $T\boldsymbol{x} = A\boldsymbol{x}$ where

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

and let

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Show that $\boldsymbol{v}_1, \boldsymbol{v}_2$ are eigenvectors of T. What are the corresponding eigenvalues? Find an ordered basis γ for $\mathbb{R}^2$ such that $[T]_\gamma$ is diagonal. Calculate $[T]_\gamma$.

Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V and suppose that

$$[T]_\gamma = D = \begin{bmatrix} d_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{nn} \end{bmatrix}$$

is a diagonal matrix for some ordered basis $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Since γ is a basis we must have $\boldsymbol{v}_i \neq \boldsymbol{0}$ for all i in $1, \ldots, n$. Furthermore, for any i in $1, \ldots, n$ we have that

$$[T\boldsymbol{v}_i]_\gamma = [T]_\gamma[\boldsymbol{v}_i]_\gamma = [T]_\gamma \boldsymbol{e}_i = d_{ii}\boldsymbol{e}_i \tag{8.2}$$

Applying ϕ_γ^{-1} to both sides of equation (8.2) gives

$$T\boldsymbol{v}_i = d_{ii}\phi_\gamma^{-1}(\boldsymbol{e}_i) = d_{ii}\boldsymbol{v}_i \tag{8.3}$$

By equation (8.3) we have for all i in $1, \ldots, n$ that $\boldsymbol{v}_i$ is an eigenvector of T with corresponding eigenvalue d_{ii}. We have justified the following lemma.

Lemma 8.1.1

Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V and suppose that $[T]_\gamma$ is a diagonal matrix for some ordered basis $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Then γ is an ordered basis for V consisting of eigenvectors of T.

Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V and suppose that $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is an ordered basis for V consisting of eigenvectors of T. Since γ contains eigenvectors of T we have for all i in $1, \ldots, n$ that $T\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$ for some $\lambda_i \in \mathbb{R}$. Since

$$\begin{aligned} [T]_\gamma &= \begin{bmatrix} [T\boldsymbol{v}_1]_\gamma & [T\boldsymbol{v}_2]_\gamma & \cdots & [T\boldsymbol{v}_n]_\gamma \end{bmatrix} \\ &= \begin{bmatrix} [\lambda_1\boldsymbol{v}_1]_\gamma & [\lambda_2\boldsymbol{v}_2]_\gamma & \cdots & [\lambda_n\boldsymbol{v}_n]_\gamma \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1[\boldsymbol{v}_1]_\gamma & \lambda_2[\boldsymbol{v}_2]_\gamma & \cdots & \lambda_n[\boldsymbol{v}_n]_\gamma \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_2 & \cdots & \lambda_n\boldsymbol{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \\ &= D \end{aligned}$$

we have that T is diagonalizable. Combining this result with lemma 8.1.1 gives us the following result.

Theorem 8.1

A linear operator $T : V \to V$ on a finite-dimensional vector space V is diagonalizable if and only if there exists an ordered basis γ for V consisting of eigenvectors of T. Furthermore, if $\gamma = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is an ordered basis for V consisting of eigenvectors of T, then

$$[T]_\gamma = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

where for all i in $1, \ldots, n$ we have that $T\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i$.

Definition 8.3

Let $A \in \mathbb{R}^{n\times n}$ and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $T\boldsymbol{x} = A\boldsymbol{x}$. Then A is called *diagonalizable* if the linear operator T is diagonalizable.

Definition 8.4

Let $A \in \mathbb{R}^{n\times n}$. A scalar $\lambda \in \mathbb{R}$ is called an *eigenvalue* of A if there exists a *nonzero* vector $\boldsymbol{x} \in \mathbb{R}^n$, called an *eigenvector* of A, such that

$$A\boldsymbol{x} = \lambda \boldsymbol{x} \tag{8.4}$$

Exercise 8.2

Let $A \in \mathbb{R}^{n\times n}$. Prove that A is diagonalizable if and only if there exists a basis for $\mathbb{R}^n$ consisting of eigenvectors of A.

Exercise 8.3

Let $A \in \mathbb{R}^{n\times n}$. Prove that A is diagonalizable if and only if there exists an invertible matrix $Q \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that

$$A = QDQ^{-1}$$

or equivalently

$$D = Q^{-1}AQ$$

Let T be a linear operator on a finite-dimensional vector space V with ordered basis β. Suppose that $\boldsymbol{v} \in V$ is an eigenvector of T with corresponding eigenvalue $\lambda \in \mathbb{R}$. Then we have that

$$T\boldsymbol{v} = \lambda\boldsymbol{v}, \qquad \boldsymbol{v} \neq \boldsymbol{0}$$

Applying ϕ_β to both sides gives

$$[T\boldsymbol{v}]_\beta = [\lambda\boldsymbol{v}]_\beta, \qquad \boldsymbol{v} \neq \boldsymbol{0}$$

and hence

$$[T]_\beta[\boldsymbol{v}]_\beta = \lambda[\boldsymbol{v}]_\beta, \qquad [\boldsymbol{v}]_\beta \neq \boldsymbol{0}$$

where $[\boldsymbol{v}]_\beta \neq \boldsymbol{0}$ since ϕ_β is one-to-one and $\boldsymbol{v} \neq \boldsymbol{0}$. Thus $[\boldsymbol{v}]_\beta \in \mathbb{R}^n$ is an eigenvector of $[T]_\beta$ with corresponding eigenvalue $\lambda \in \mathbb{R}$. We have justified the first part of the following theorem. We leave the remainder of the proof as an exercise.

Theorem 8.2

Let T be a linear operator on a finite-dimensional vector space V with ordered basis β. Then λ is an eigenvalue of T if and only if λ is an eigenvalue of $[T]_\beta$. Furthermore,

(1) If $\boldsymbol{v} \in V$ is an eigenvector of T with corresponding eigenvalue λ, then $\phi_\beta(\boldsymbol{v}) \in \mathbb{R}^n$ is an eigenvector of $[T]_\beta$ with corresponding eigenvalue λ.

(2) If $\boldsymbol{x} \in \mathbb{R}^n$ is an eigenvector of $[T]_\beta$ with corresponding eigenvalue λ, then $\phi_\beta^{-1}(\boldsymbol{x}) \in V$ is an eigenvector of T with corresponding eigenvalue λ.

Exercise 8.4

Prove theorem 8.2.

Exercise 8.5

Let T be a linear operator on a finite-dimensional vector space V with ordered basis β.

(1) Prove that if γ is an ordered basis for V consisting of eigenvectors of T then $\phi_\beta(\gamma)$ is an ordered basis for $\mathbb{R}^n$ consisting of eigenvectors of $[T]_\beta$.

(2) Prove that if α is an ordered basis for $\mathbb{R}^n$ consisting of eigenvectors of $[T]_\beta$, then $\phi_\beta^{-1}(\alpha)$ is an ordered basis for V consisting of eigenvectors of T.

The next theorem allows us to focus on the problem of testing to see if a matrix is diagonalizable. We leave the proof as an exercise.

Theorem 8.3

Let T be a linear operator on a finite-dimensional vector space V with ordered basis β. Then T is diagonalizable if and only if $[T]_\beta$ is diagonalizable.

Exercise 8.6

Prove theorem 8.3.

Let T be the linear operator on $P_2(\mathbb{R})$ defined by

$$(T\boldsymbol{f})(t) = f(t) + (t+1)f'(t)$$

and let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. Since

$$T1 = 1, \qquad Tt = 1 + 2t, \qquad Tt^2 = 2t + 3t^2$$

we have that

$$[T]_\beta = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

It is easy to check that

$$\alpha = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$$

where

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

is an ordered basis for $\mathbb{R}^3$ consisting of eigenvectors of $[T]_\beta$. We set $Q = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, and calculate

$$QDQ^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = [T]_\beta$$

which verifies that $[T]_\beta$ is diagonalizable by exercise 8.3. Using exercise 8.5 we deduce that

$$\gamma = \phi_\beta^{-1}(\alpha) = \{1, 1+t, 1+2t+t^2\}$$

is an ordered basis for $P_2(\mathbb{R})$ consisting of eigenvectors of T. Finally we note that

$$T1 = (1)1, \qquad T(1+t) = 2 + 2t = 2(1+t), \qquad T(1+2t+t^2) = 3 + 6t + 3t^2 = 3(1+2t+t^2)$$

$$[T]_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

which verifies that T is diagonalizable.

8.2. Determinants

Let V be a vector space such that $\dim(V) = n$. Let $T : V \to V$ be a linear operator and β be an ordered basis for V. By theorem 8.3, the linear operator T is diagonalizable if and only $[T]_\beta$ is diagonalizable. Furthermore, $[T]_\beta$ is diagonalizable if and only if there exists an ordered basis for $\mathbb{R}^n$ consisting of eigenvectors of $[T]_\beta$. We will thus need a procedure for finding an ordered basis for $\mathbb{R}^n$ consisting of eigenvectors of $[T]_\beta$ if such a basis exists. To look for a basis of eigenvectors, we must first find all eigenvalues of the matrix $[T]_\beta$.

To find eigenvalues of a matrix $A \in \mathbb{R}^{n\times n}$ we look for $\lambda \in \mathbb{R}$ and $\boldsymbol{x} \neq \mathbf{0}$ such that

$$A\boldsymbol{x} = \lambda\boldsymbol{x}$$

or equivalently

$$(A - \lambda I_n)\boldsymbol{x} = \mathbf{0}, \qquad \boldsymbol{x} \neq \mathbf{0} \tag{8.5}$$

Thus, we look for $\lambda \in \mathbb{R}$ such that $N(A - \lambda I_n)$ contains a nonzero vector or equivalently $A - \lambda I_n$ is not invertible by exercise 5.19. To find $\lambda \in \mathbb{R}$ such that $A - \lambda I_n$ is not invertible, we will use the *determinant*.

Definition 8.5

Let $A \in \mathbb{R}^{n\times n}$. The *determinant* of A is a scalar, denoted by $\det(A)$ or $|A|$, and is defined recursively as follows.

(1) If $A \in \mathbb{R}^{1\times 1}$, then $\det(A) = a_{11}$, the single entry of A.
(2) If $A \in \mathbb{R}^{n\times n}$ for $n > 1$, then

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(\tilde{a}_{1j})$$

where $\tilde{a}_{1j}$ is the $(n-1) \times (n-1)$ matrix obtained by deleting row 1 and column j from the matrix A.

Exercise 8.7

Suppose A is a 2×2 matrix. Show that $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

Exercise 8.8

Calculate $\det(A)$ where $A = \begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix}$.

Exercise 8.9

Calculate $\det(A)$ where $A = \begin{bmatrix} 1 & 1 & 5 \\ 1 & -4 & -1 \\ 0 & -3 & 1 \end{bmatrix}$.

The next result says that the determinant can be expanded along any row or column. We omit the proof.

Theorem 8.4

Let $A \in \mathbb{R}^{n\times n}$. Then for any $1 \le i \le n$ and any $1 \le j \le n$ we have

$$\det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(\tilde{a}_{ij}) = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(\tilde{a}_{ij})$$

where $\tilde{a}_{ij}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting row i and column j from A.

Consider the matrix

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

To calculate $\det(A)$ efficiently, we use theorem 8.4 and expand along the second column to get

$$\det(A) = (-1)^{1+2}(1)\det\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right) = (-1)(-2) = 2$$

A matrix $A \in \mathbb{R}^{n\times n}$ is called *upper triangular* if all entries lying below the diagonal entries are zero (i.e. $a_{ij} = 0$ whenever $i > j$). A matrix $A \in \mathbb{R}^{n\times n}$ is called *lower triangular* if all entries lying above the diagonal entries are zero (i.e. $a_{ij} = 0$ whenever $i < j$). A matrix that is both upper triangular and lower triangular is diagonal. The next result says that it is easy to calculate the determinant for such matrices. We leave the proof as an exercise.

Lemma 8.2.1

Let $A \in \mathbb{R}^{n\times n}$ be an upper triangular, lower triangular, or diagonal matrix. Then

$$\det(A) = a_{11}a_{22}\cdots a_{nn}$$

Exercise 8.10

Prove lemma 8.2.1.

Exercise 8.11

Calculate $\det(A)$ where $A = \begin{bmatrix} -3 & 1 & 2 \\ 0 & 4 & 5 \\ 0 & 0 & -6 \end{bmatrix}$.

One easy consequence of lemma 8.2.1 is that for any $n > 0$ we have that

$$\det(I_n) = 1 \tag{8.6}$$

Consider the matrix

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

Expanding along the second column or third row will be the most efficient due to the location of the 0. However, many calculations will still be required. The following theorem, which we state without proof, allows us to use elementary row operations to further simplify the calculation of determinants.

Theorem 8.5

Let $A \in \mathbb{R}^{n\times n}$.

(1) If B is a matrix obtained by interchanging any two rows or interchanging any two columns of A, then $\det(B) = -\det(A)$.

(2) If B is a matrix obtained by multiplying a row or a column of A by a scalar k, then $\det(B) = k\det(A)$.

(3) If B is a matrix obtained from A by adding a multiple of row i to row j or a multiple of column i to column j for $i \neq j$, then $\det(B) = \det(A)$.

Again consider the matrix

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

Adding -1 times row 1 to row 2 and -1 times row 1 to row 4 gives

$$B = \begin{bmatrix} 2 & 1 & 1 & 5 \\ -1 & 0 & -5 & -6 \\ 2 & 0 & -3 & 1 \\ 1 & 0 & 0 & -3 \end{bmatrix}$$

By theorem 8.5 we have that $\det(A) = \det(B)$. Next we expand along the second column of B using theorem 8.4 to get

$$\det(A) = \det(B) = (-1)^{1+2}(1)\det\left(\begin{bmatrix} -1 & -5 & -6 \\ 2 & -3 & 1 \\ 1 & 0 & -3 \end{bmatrix}\right)$$

Let $C = \begin{bmatrix} -1 & -5 & -6 \\ 2 & -3 & 1 \\ 1 & 0 & -3 \end{bmatrix}$. Adding 3 times the first column to the third column gives

$$D = \begin{bmatrix} -1 & -5 & -9 \\ 2 & -3 & 7 \\ 1 & 0 & 0 \end{bmatrix}$$

By theorem 8.5 we have that $\det(C) = \det(D)$. Finally, we expand along the third row of D using theorem 8.4 to get

$$\det(A) = \det(B) = -1\det(C) = -1\det(D) = (-1)(-1)^{3+1}(1)\det\left(\begin{bmatrix} -5 & -9 \\ -3 & 7 \end{bmatrix}\right) = 62$$

Exercise 8.12

Consider the matrix $A = \begin{bmatrix} 1 & 0 & -2 & 3 \\ -3 & 1 & 1 & 2 \\ 0 & 4 & -1 & 1 \\ 2 & 3 & 0 & 1 \end{bmatrix}$. Write $\det(A)$ as the determinant of a single 3×3 matrix by expanding the determinant along the first column after first applying certain row operations. Calculate the determinant of A.

Let $A \in \mathbb{R}^{n\times n}$ with column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and suppose that A is not invertible. By exercise 5.19 we have that $N(A) \neq \{\mathbf{0}\}$. Let $\boldsymbol{x}$ be a nonzero vector in $N(A)$. Then we have that

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \mathbf{0}$$

Since $\boldsymbol{x} \neq \mathbf{0}$, we have $x_i \neq 0$ for some i in $1, \ldots, n$ and thus

$$\boldsymbol{a}_i = \frac{-1}{x_i}\left(x_1\boldsymbol{a}_1 + \cdots + x_{i-1}\boldsymbol{a}_{i-1} + x_{i+1}\boldsymbol{a}_{i+1} + \cdots + x_n\boldsymbol{a}_n\right) \tag{8.7}$$

Equation (8.7) shows that we can transform the i^{th} column of A into $\mathbf{0}$ by adding multiples of the other $n-1$ column vectors without changing the determinant by theorem 8.5. Finally, expanding along the i^{th} column shows that $\det(A) = 0$. We summarize this result in the following lemma.

Lemma 8.2.2

Let $A \in \mathbb{R}^{n\times n}$. If A is not invertible, then $\det(A) = 0$.

The next lemma says that the determinant is multipliciative. We omit the proof.

Lemma 8.2.3

If $A, B \in \mathbb{R}^{n\times n}$, then

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

Suppose $A \in \mathbb{R}^{n\times n}$ is invertible. Then we have

$$AA^{-1} = I_n$$

By lemma 8.2.3 and equation (8.6) we have that

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$$

and thus $\det(A) \neq 0$. Combining this result with lemma 8.2.2 gives us the following result.

Theorem 8.6

Suppose $A \in \mathbb{R}^{n\times n}$. Then A is invertible if and only if $\det(A) \neq 0$.

The following corollary gives us our primary tool for finding all eigenvalues of a matrix A using the determinant. We leave the proof as an exercise.

Corollary 8.2.1

Let $A \in \mathbb{R}^{n\times n}$. Then the following statements are equivalent for all $\lambda \in \mathbb{R}$.

(1) λ is an eigenvalue of A.
(2) $N(A - \lambda I_n) \neq \{\mathbf{0}\}$.
(3) $A - \lambda I_n$ is not invertible.
(4) $\det(A - \lambda I_n) = 0$.

Exercise 8.13

Prove corollary 8.2.1.

Let $A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$. To find all eigenvalues of A we solve

$$\det(A - \lambda I_2) = \det\left(\begin{bmatrix} 1-\lambda & 1 \\ 4 & 1-\lambda \end{bmatrix}\right) = (1-\lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = 0$$

Since

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

the eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$ by corollary 8.2.1.

Let $A = \begin{bmatrix} 2 & 4 & 6 \\ 0 & 2 & 0 \\ 0 & 2 & 4 \end{bmatrix}$. To find all eigenvalues of A we solve

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & 4 & 6 \\ 0 & 2-\lambda & 0 \\ 0 & 2 & 4-\lambda \end{bmatrix}\right) = (2-\lambda)^2(4-\lambda)$$

Thus the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 4$ by corollary 8.2.1.

Let $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$. To find all eigenvalues of A we solve

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

Thus the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 0$ by corollary 8.2.1.

Exercise 8.14

Let $A \in \mathbb{R}^{3\times 3}$ be given by

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

Use corollary 8.2.1 to find all eigenvalues of A.

Exercise 8.15

Let $A \in \mathbb{R}^{4\times 4}$ be given by

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

Use corollary 8.2.1 to find all eigenvalues of A.

The next exercise gives us an easy way to find the eigenvalues of matrix that is upper triangular, lower triangular, or diagonal.

Exercise 8.16

Let $A \in \mathbb{R}^{n\times n}$ be an upper triangular, lower triangular, or diagonal matrix. Prove that A has eigenvalues

$$\lambda_1 = a_{11}, \qquad \lambda_2 = a_{22}, \qquad \cdots \qquad \lambda_n = a_{nn}$$

Consider the matrix

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

Since A is upper triangular, the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = -1$ by exercise 8.16.

Let T be the linear operator on $P_2(\mathbb{R})$ defined by

$$(T\boldsymbol{f})(t) = f(t) + (t+1)f'(t)$$

and let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. We showed earlier that

$$[T]_\beta = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

By exercise 8.16, the eigenvalues of $[T]_\beta$ are $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$. By theorem 8.2, the linear map T also has eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$.

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by

$$(T\boldsymbol{f})(t) = f(t) - 2tf'(t) + (t^2+1)f''(t)$$

and let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. We calculate

$$T1 = 1, \qquad Tt = -t, \qquad Tt^2 = 2 - t^2$$

and thus we have

$$[T]_\beta = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

By exercise 8.16, the eigenvalues of $[T]_\beta$ are $\lambda_1 = 1$ and $\lambda_2 = -1$. By theorem 8.2, the linear map T also has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$.

Exercise 8.17

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by

$$(T\boldsymbol{f})(t) = tf'(t) + f(2)t + f(3)$$

Find the eigenvalues of T.

Definition 8.6

Let $A, B \in \mathbb{R}^{n\times n}$. We say that A is *similar* to B if $A = CBC^{-1}$ for some invertible matrix $C \in \mathbb{R}^{n\times n}$.

Note that by exercise 8.3, a matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if A is similar to a diagonal matrix $D \in \mathbb{R}^{n\times n}$.

Exercise 8.18

Let $A, B \in \mathbb{R}^{n\times n}$.

(1) Prove that if A is similar to B, then B is similar to A.
(2) Prove that if A is similar to B, then $\operatorname{rank}(A) = \operatorname{rank}(B)$.
(3) Prove that if A is similar to B, then $\det(A) = \det(B)$.
(4) Prove that if A is similar to B, then $A - \lambda I_n$ is similar to $B - \lambda I_n$. Use corollary 8.2.1 to conclude that if A is similar to B, then A and B have the same eigenvalues.

8.3. Eigenspaces

Now that we know how to find all eigenvalues of a matrix $A \in \mathbb{R}^{n\times n}$ using the determinant, the next step in finding a basis for $\mathbb{R}^n$ consisting of eigenvectors of A is to find the corresponding eigenvectors for each eigenvalue.

Let $A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$. We showed earlier that the eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$. To find an eigenvector of A with corresponding eigenvalue $\lambda_1 = 3$, we look for a nonzero $\boldsymbol{x} \in \mathbb{R}^2$ such that

$$A\boldsymbol{x} = 3\boldsymbol{x}$$

or equivalently

$$(A - 3I_2)\boldsymbol{x} = \mathbf{0}$$

Since $A - 3I_2 = \begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix}$ we want to solve

$$\begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

for some $\boldsymbol{x} \neq \mathbf{0}$. Since

$$1 \begin{bmatrix} -2 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \mathbf{0}$$

one eigenvector is

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Another eigenvector is $\boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. How do we describe all possible eigenvectors with corresponding eigenvalue $\lambda_1 = 3$? Since $\mathbf{0}$ is not an eigenvector, the set of eigenvectors with corresponding eigenvalue $\lambda_1 = 3$ is *not* a subspace of $\mathbb{R}^2$. However, the set of eigenvectors plus $\mathbf{0}$ is a subspace of $\mathbb{R}^2$.

Definition 8.7

Let $A \in \mathbb{R}^{n\times n}$ and let λ be an eigenvalue of A. Define

$$E_\lambda = N(A - \lambda I_n)$$

The set E_λ is called the *eigenspace* of A corresponding to the eigenvalue λ.

Note that E_λ is a subspace of $\mathbb{R}^n$ consisting of the zero vector and the eigenvectors of A corresponding to the eigenvalue λ. Furthermore, by the rank-nullity theorem we have

$$\dim(E_\lambda) = n - \operatorname{rank}(A - \lambda I_n) \tag{8.8}$$

Let $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$. We showed earlier that the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 0$. By the rank-nullity theorem we have

$$\dim(E_{\lambda_1}) = 4 - \operatorname{rank}(A - 2I_4) = 4 - \operatorname{rank}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}\right) = 4 - 1 = 3$$

Since $\dim(E_{\lambda_1}) = 3$, we can find a basis for E_{λ_1} by finding a linearly independent set of 3 vectors in $E_{\lambda_1} = N(A - 2I_4)$ by solving

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

We see by inspection that

$$\alpha_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a linearly independent set of 3 solutions and hence α_1 is a basis for E_{λ_1}. By the rank-nullity theorem we have

$$\dim(E_{\lambda_2}) = 4 - \operatorname{rank}(A - 0I_4) = 4 - \operatorname{rank}\left(\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \right) = 4 - 3 = 1$$

Since $\dim(E_{\lambda_2}) = 1$, we can find a basis for E_{λ_2} by finding a single nonzero vector in $E_{\lambda_2} = N(A - 2I_4)$ by solving

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

We see by inspection that

$$\alpha_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a basis for E_{λ_2}. To find a basis for $\mathbb{R}^4$ consisting of eigenvectors of A, we need a linearly independent set of 4 eigenvectors of A. It is easy to verify that

$$\alpha = \alpha_1 \cup \alpha_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^4$ consisting of eigenvectors of A and thus A is diagonalizable by exercise 8.2. Note that if we set

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

then

$$QDQ^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 & -0.5 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = A$$

and we have verified that A is diagonalizable.

Exercise 8.19

Let $A \in \mathbb{R}^{2\times 2}$ be given by

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

Find the eigenvalues of A. For each eigenvalue λ_i calculate $\dim(E_{\lambda_i})$ and find a basis for E_{λ_i}. Find a basis α for $\mathbb{R}^2$ consisting of eigenvectors of A. Find an invertible matrix $Q \in \mathbb{R}^{2\times 2}$ and a diagonal matrix $D \in \mathbb{R}^{2\times 2}$ such that $A = QDQ^{-1}$.

Exercise 8.20

Let $A \in \mathbb{R}^{3\times 3}$ be given by

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

Find the eigenvalues of A. For each eigenvalue λ_i calculate $\dim(E_{\lambda_i})$ and find a basis for E_{λ_i}. Find a basis α for $\mathbb{R}^3$ consisting of eigenvectors of A. Find an invertible matrix $Q \in \mathbb{R}^{3\times 3}$ and a diagonal matrix $D \in \mathbb{R}^{3\times 3}$ such that $A = QDQ^{-1}$.

Exercise 8.21

Let $A \in \mathbb{R}^{4\times 4}$ be given by

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

Find the eigenvalues of A. For each eigenvalue λ_i calculate $\dim(E_{\lambda_i})$ and find a basis for E_{λ_i}. Find a basis α for $\mathbb{R}^4$ consisting of eigenvectors of A. Find an invertible matrix $Q \in \mathbb{R}^{4\times 4}$ and a diagonal matrix $D \in \mathbb{R}^{4\times 4}$ such that $A = QDQ^{-1}$.

Exercise 8.22

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has *distinct* eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$. Prove that $E_{\lambda_1} \cap E_{\lambda_2} = \{\mathbf{0}\}$.

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by

$$(T\boldsymbol{f})(t) = f(t) - 2tf'(t) + (t^2+1)f''(t)$$

Let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. We showed earlier that

$$[T]_\beta = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

and the eigenvalues of $[T]_\beta$ are $\lambda_1 = 1$ and $\lambda_2 = -1$. By the rank-nullity theorem we have

$$\dim(E_{\lambda_1}) = 3 - \operatorname{rank}(A - 1I_3) = 3 - \operatorname{rank}\left(\begin{bmatrix} 0 & 0 & 2 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}\right) = 3 - 2 = 1$$

Since $\dim(E_{\lambda_1}) = 1$, we can find a basis for E_{λ_1} by finding a nonzero vector in $E_{\lambda_1} = N(A - 1I_3)$ by solving

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}\boldsymbol{x} = \mathbf{0}$$

We see by inspection that

$$\alpha_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

a basis for E_{λ_1}. By the rank-nullity theorem we have

$$\dim(E_{\lambda_2}) = 3 - \operatorname{rank}(A + 1I_3) = 3 - \operatorname{rank}\left(\begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) = 3 - 1 = 2$$

Since $\dim(E_{\lambda_2}) = 2$, we can find a basis for E_{λ_2} by finding a linearly independent set of 2 vectors in $E_{\lambda_2} = N(A + I_3)$ by solving

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

We see by inspection that

$$\alpha_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a linearly independent set of 2 solutions and hence α_1 is a basis for E_{λ_2}. To find a basis for $\mathbb{R}^3$ consisting of eigenvectors of A, we need a linearly independent set of 3 eigenvectors of A. It is easy to verify that

$$\alpha = \alpha_1 \cup \alpha_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^3$ consisting of eigenvectors of A. Using exercise 8.5 we deduce that

$$\gamma = \phi_\beta^{-1}(\alpha) = \{1, t, 1 - t^2\}$$

is an ordered basis for $P_2(\mathbb{R})$ consisting of eigenvectors of T. As noted earlier we have

$$[T]_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

which verifies that T is a diagonalizable operator.

Exercise 8.23

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by

$$(T\boldsymbol{f})(t) = f(t) + (t + 1)f'(t)$$

Find an ordered basis γ for $P_2(\mathbb{R})$ such that $[T]_\gamma$ is a diagonal matrix. Verify your answer by calculating $[T]_\gamma$.

Exercise 8.24

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by

$$(T\boldsymbol{f})(t) = tf'(t) + f(2)t + f(3)$$

Find an ordered basis γ for $P_2(\mathbb{R})$ such that $[T]_\gamma$ is a diagonal matrix. Verify your answer by calculating $[T]_\gamma$.

In previous examples we took the union of eigenspace bases to find a basis for $\mathbb{R}^n$ consisting of eigenvectors of a matrix A. In general, the union of linearly independent sets does not have to be linearly independent. For example, the set $\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\} \cup \left\{\begin{bmatrix}2\\2\end{bmatrix}\right\}$ is linearly dependent. It turns out that the union of linearly independent sets from different eigenspaces is always a linearly independent set.

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has *distinct* eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a linearly independent set of n vectors in E_{λ_1} and $T = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}$ be a linearly independent set of m vectors in E_{λ_2}. Suppose that

$$c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n + d_1\boldsymbol{w}_1 + \cdots + d_m\boldsymbol{w}_m = \mathbf{0} \tag{8.9}$$

We next multiply both sides by the matrix $A - \lambda_1 I_n$ to get

$$\begin{aligned}
\mathbf{0} &= (A-\lambda_1 I_n)\mathbf{0}\\
&= (A-\lambda_1 I_n)(c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n + d_1\boldsymbol{w}_1 + \cdots + d_m\boldsymbol{w}_m)\\
&= c_1(A-\lambda_1 I_n)\boldsymbol{v}_1 + \cdots + c_n(A-\lambda_1 I_n)\boldsymbol{v}_n + d_1(A-\lambda_1 I_n)\boldsymbol{w}_1 + \cdots + d_m(A-\lambda_1 I_n)\boldsymbol{w}_m\\
&= c_1(A\boldsymbol{v}_1 - \lambda_1\boldsymbol{v}_1) + \cdots + c_n(A\boldsymbol{v}_n - \lambda_1\boldsymbol{v}_n) + d_1(A\boldsymbol{w}_1 - \lambda_1\boldsymbol{w}_1) + \cdots + d_m(A\boldsymbol{w}_m - \lambda_1\boldsymbol{w}_m)\\
&= c_1(\lambda_1\boldsymbol{v}_1 - \lambda_1\boldsymbol{v}_1) + \cdots + c_n(\lambda_1\boldsymbol{v}_n - \lambda_1\boldsymbol{v}_n) + d_1(\lambda_2\boldsymbol{w}_1 - \lambda_1\boldsymbol{w}_1) + \cdots + d_m(\lambda_2\boldsymbol{w}_m - \lambda_1\boldsymbol{w}_m)\\
&= (\lambda_2 - \lambda_1)(d_1\boldsymbol{w}_1 + \cdots + d_m\boldsymbol{w}_m)
\end{aligned}$$

Since $\lambda_2 - \lambda_1 \neq 0$, we must have

$$d_1\boldsymbol{w}_1 + \cdots + d_m\boldsymbol{w}_m = \mathbf{0}$$

and since T is linearly independent

$$d_1 = \cdots = d_m = 0 \tag{8.10}$$

Plugging equation (8.10) into equation (8.9) and using the fact that S is linearly independent gives

$$c_1 = \cdots = c_n = 0 \tag{8.11}$$

and we have shown that $S \cup T$ is a linearly independent set of $n+m$ vectors. We have justified the following theorem for the special case $k=2$. We leave the proof of the general case as an exercise.

Theorem 8.7

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has k *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. For each $i = 1, 2\ldots, k$, let S_i be a linearly independent subset of E_{λ_i} containing n_i vectors. Then $S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a linearly independent set in $\mathbb{R}^n$ containing $n_1 + n_2 + \cdots + n_k$ vectors.

Exercise 8.25

Prove theorem 8.7. Hint: Use mathematical induction on k.

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has k *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. Further suppose that

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k}) = n$$

For i in $1, \ldots, k$, let α_i be a basis for E_{λ_i} containing $\dim(E_{\lambda_i})$ vectors. By theorem 8.7 and corollary 5.3.2 we have that

$$\alpha = \alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_k$$

is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A and thus A is diagonalizable by exercise 8.2. We have justified the following result.

Corollary 8.3.1

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has k *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. If

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k}) = n$$

then A is diagonalizable.

The matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ has eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$. We calculate

$$\dim(E_{\lambda_1}) = 3 - \text{rank}\left(\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}\right) = 3 - 2 = 1$$

$$\dim(E_{\lambda_2}) = 3 - \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right) = 3 - 1 = 2$$

Since $\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) = 3$, we have that A is diagonalizable by corollary 8.3.1.

Exercise 8.26

Let $A \in \mathbb{R}^{n \times n}$ and suppose A has k *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. Prove that

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k}) \leq n$$

Exercise 8.27

Let $A \in \mathbb{R}^{n \times n}$ and suppose A has n *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Prove that A is diagonalizable.

The matrix $A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{bmatrix}$ has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 2$. We calculate

$$\dim(E_{\lambda_1}) = 4 - \text{rank}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\right) = 4 - 3 = 1$$

$$\dim(E_{\lambda_2}) = 4 - \text{rank}\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}\right) = 4 - 2 = 2$$

Since $\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) = 3 < 4$, corollary 8.3.1 does not apply. If we take the union of a basis for E_{λ_1} and a basis for E_{λ_2} we will have a linearly indpendent set of 3 eigenvectors. However, a basis for $\mathbb{R}^4$ requires 4 vectors. We suspect that A is not diagonalizble but we need another result to justify it.

Let $A \in \mathbb{R}^{n \times n}$ and suppose that A is diagonalizable. Then by exercise 8.2 there exists a basis α for $\mathbb{R}^n$ containing eigenvectors of A. Suppose A has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. For $1 \leq i \leq k$, let

$$\alpha_i = \alpha \cap E_{\lambda_i}$$

and n_i be the number of vectors in α_i. Note that $n \leq n_1 + n_2 + \cdots + n_k$ since each eigenvector must have a corresponding eigenvalue. Since each α_i is linearly independent by exercise 4.13 we must have $n_i \leq \dim(E_{\lambda_i})$ for all $1 \leq i \leq k$ by theorem 5.2. Combining these results gives

$$n \leq n_1 + n_2 + \cdots + n_k \leq \dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k})$$

and hence

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k}) = n$$

by exercise 8.26. This result combined with corollary 8.3.1 justifies the following result which gives us our main test for diagonalizability of a matrix.

Theorem 8.8

Let $A \in \mathbb{R}^{n\times n}$ and suppose A has k *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. Then A is diagonalizable if and only if

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_k}) = n$$

Consider again the matrix $A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{bmatrix}$ with eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 2$. Since $\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) = 3 < 4$, the matrix A is not diagonalizable by theorem 8.8.

Exercise 8.28

Let $A = \begin{bmatrix} -1 & 0 & 1 \\ 3 & 0 & -3 \\ 1 & 0 & -1 \end{bmatrix}$. Is A diagonalizable? Explain.

Exercise 8.29

Let $A = \begin{bmatrix} 2 & 4 & 6 \\ 0 & 2 & 0 \\ 0 & 2 & 4 \end{bmatrix}$. Is A diagonalizable? Explain.

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by $(T\boldsymbol{f})(t) = f(t) - 2f'(t)$. Let $\beta = \{1, t, t^2\}$ be the standard ordered basis for $P_2(\mathbb{R})$. We calculate

$$T1 = 1, \qquad Tt = t - 2, \qquad Tt^2 = t^2 - 4t$$

$$[T]_\beta = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$$

$[T]_\beta$ has the single eigenvalue $\lambda = 1$. We calculate

$$\dim(E_\lambda) = 3 - \operatorname{rank}\left(\begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{bmatrix}\right) = 3 - 2 = 1$$

Since $\dim(E_\lambda) = 1 \neq 3$ we have that $[T]_\beta$ is not diagonalizable by theorem 8.8 and hence T is not diagonalizable by theorem 8.3.

Exercise 8.30

Consider the linear operator $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ given by $(T\boldsymbol{f})(t) = f(t) + f'(t) + f''(t)$. Is T diagonalizable? Explain.

Exercise 8.31

Find a matrix $A \in \mathbb{R}^{2\times 2}$ such that

(1) A is invertible but not diagonalizable.
(2) A is invertible and diagonalizable.
(3) A is not invertible but is diagonalizable.
(4) A is not invertible and not diagonalizable.

CHAPTER 9

Inner Product Spaces

9.1. Inner Products

Definition 9.1

Definition: An *inner product* on a vector space V is a function that takes two vectors $\boldsymbol{x}, \boldsymbol{y} \in V$ and returns a scalar in $\mathbb{R}$ denoted $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ such that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and all c in $\mathbb{R}$ the following axioms hold.

(1) $\langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{z}, \boldsymbol{y} \rangle$
(2) $\langle c\boldsymbol{x}, \boldsymbol{y} \rangle = c\langle \boldsymbol{x}, \boldsymbol{y} \rangle$
(3) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
(4) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$ if $\boldsymbol{x} \neq \boldsymbol{0}$

Definition 9.2

A vector space V with inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is called an *inner product space*.

For $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$ we define $\boldsymbol{x} \cdot \boldsymbol{y}$ (called the *dot product*) as

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \tag{9.1}$$

Note that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have that

$$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y} \tag{9.2}$$

Exercise 9.1

Show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot \boldsymbol{y}$ is an inner product on $\mathbb{R}^n$.

We call $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot \boldsymbol{y}$ the *standard inner product* on $\mathbb{R}^n$.

Definition 9.3

A matrix $A \in \mathbb{R}^{n \times n}$ is called *symmetric positive definite* if $A^T = A$ and $\boldsymbol{x}^T A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$.

Suppose that $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix and for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ define

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot A\boldsymbol{y} \tag{9.3}$$

We will show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is an inner product on $\mathbb{R}^n$.

(1) For $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$ we note that

$$\langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} \rangle = (\boldsymbol{x} + \boldsymbol{z}) \cdot A\boldsymbol{y} = \boldsymbol{x} \cdot A\boldsymbol{y} + \boldsymbol{z} \cdot A\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{z}, \boldsymbol{y} \rangle$$

(2) For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we calculate

$$\langle c\boldsymbol{x}, \boldsymbol{y}\rangle = (c\boldsymbol{x}) \cdot A\boldsymbol{y} = c(\boldsymbol{x} \cdot A\boldsymbol{y}) = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$$

(3) For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have that

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x} \cdot A\boldsymbol{y} = A\boldsymbol{y} \cdot \boldsymbol{x} = (A\boldsymbol{y})^T\boldsymbol{x} = \boldsymbol{y}^T A^T \boldsymbol{x} = \boldsymbol{y}^T A\boldsymbol{x} = \boldsymbol{y} \cdot A\boldsymbol{x} = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$$

(4) Let $\boldsymbol{x} \in \mathbb{R}^n$ and suppose that $\boldsymbol{x} \neq \boldsymbol{0}$. We calculate

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \boldsymbol{x} \cdot A\boldsymbol{x} = \boldsymbol{x}^T A\boldsymbol{x} > 0$$

since the matrix A is *positive definite*.

Definition 9.4

Let $C([a,b])$ be the subset of $\mathcal{F}([a,b],\mathbb{R})$ of all functions that are continuous for all $t \in [a,b]$.

Exercise 9.2

Prove that $C([a,b])$ is a subspace of $\mathcal{F}([a,b],\mathbb{R})$.

Exercise 9.3

For $\boldsymbol{f}, \boldsymbol{g} \in C([a,b])$ define $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_a^b f(t)g(t)\,dt$. Prove that $\langle x, y\rangle$ is an inner product on $C([a,b])$.

The following result gives some basic inner product results that follow from the axioms. We leave the proof as an exercise.

Lemma 9.1.1

Let $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ be an inner product on V. For $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and c in $\mathbb{R}$ we have

(1) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle$.
(2) $\langle \boldsymbol{x}, c\boldsymbol{y}\rangle = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$.
(3) $\langle \boldsymbol{x}, \boldsymbol{0}\rangle = \langle \boldsymbol{0}, \boldsymbol{x}\rangle = 0$.
(4) $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$ and $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$.
(5) If $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle$ for all $\boldsymbol{x} \in V$, then $\boldsymbol{y} = \boldsymbol{z}$.

Exercise 9.4

Prove lemma 9.1.1.

9.2. Norms and Orthogonality

Definition 9.5

Let V be an inner product space. For $\boldsymbol{x} \in V$, we define the *norm* or *length* of $\boldsymbol{x}$ by

$$\|x\| = \sqrt{\langle x, x\rangle} \tag{9.4}$$

Let $V = \mathbb{R}^n$ with the standard inner product. Then

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2} \tag{9.5}$$

Exercise 9.5

Let V be an inner product space. Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and all $c \in \mathbb{R}$ the following properties hold.

(1) $\|c\boldsymbol{x}\| = |c|\,\|\boldsymbol{x}\|$
(2) $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$

Consider a right triangle with side lengths $a, b, c \in \mathbb{R}$ where c is the length of the hypotenuse. The Pythagorean theorem for right triangles says that

$$a^2 + b^2 = c^2 \tag{9.6}$$

Now consider two vectors $\boldsymbol{x}, \boldsymbol{y}$ in an inner product space V. Equation (9.6) suggests that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal then we should have

$$\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = \|\boldsymbol{x} + \boldsymbol{y}\|^2 \tag{9.7}$$

Using properties of the inner product we calculate

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \|\boldsymbol{y}\|^2 \tag{9.8}$$

and note that in order for equation 9.7 to hold we must have $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

Definition 9.6

Let V be an inner product space. We say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

Equation (9.8) justifies the following theorem.

Theorem 9.1: Pythagorean Theorem

Let V be an inner product space. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal, then

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$$

Let V be an inner product space. By lemma 9.1.1, we have that $\boldsymbol{0}$ is orthogonal to any vector $\boldsymbol{x} \in V$ and furthermore $\boldsymbol{0}$ is the only vector that is orthogonal to itself.

Exercise 9.6

Prove that the vectors $1, t$ in $C([-1, 1])$ are orthogonal. What is the length of the vector t in $C([-1, 1])$?

Exercise 9.7

Let V be an inner product space. Consider the vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ in V and suppose that

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 0, \qquad \langle \boldsymbol{v}_1, \boldsymbol{v}_3 \rangle = 0, \qquad \langle \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle = 0$$

Prove that

$$\|\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3\|^2 = \|\boldsymbol{v}_1\|^2 + \|\boldsymbol{v}_2\|^2 + \|\boldsymbol{v}_3\|^2$$

Exercise 9.8

Let V be an inner product space. Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in V$

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$$

Exercise 9.9

Let V be a finite-dimensional inner product space with generating set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. Suppose that for some $\boldsymbol{x} \in V$ we have that $\langle \boldsymbol{x}, \boldsymbol{v}_i \rangle = 0$ for all $i \in 1, \ldots, n$. Prove that $\boldsymbol{x} = \boldsymbol{0}$. Hint: Show that $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$.

Exercise 9.10

Suppose $A \in \mathbb{R}^{m \times n}$ has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Let $\boldsymbol{x} \in \mathbb{R}^m$. Prove that $\boldsymbol{x} \in N(A^T)$ if and only if $\boldsymbol{x} \cdot \boldsymbol{a}_i = 0$ for all i in $1, \ldots, n$.

Let V be an inner product space. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in V where $\boldsymbol{y} \neq \boldsymbol{0}$. By properties of the inner product we have that

$$\langle \boldsymbol{x} - c\boldsymbol{y}, \boldsymbol{y} \rangle = 0 \Leftrightarrow c = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle}$$

Definition 9.7

Let V be an inner product space. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in V where $\boldsymbol{y} \neq \boldsymbol{0}$. The *orthogonal projection* of $\boldsymbol{x}$ onto $\boldsymbol{y}$, denoted $\operatorname{proj}_{\boldsymbol{y}} \boldsymbol{x}$ is defined by

$$\operatorname{proj}_{\boldsymbol{y}} \boldsymbol{x} = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} \boldsymbol{y} = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2} \boldsymbol{y} = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|} \left(\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \right)$$

The number $\dfrac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|}$ is called the *component* of $\boldsymbol{x}$ along $\boldsymbol{y}$ and is denoted $\operatorname{comp}_{\boldsymbol{y}} \boldsymbol{x}$.

Let $V = \mathbb{R}^2$ with the standard inner product. Consider the problem of finding the distance from the point $P = (7, 6)$ to the line through the origin in the direction of the vector $\boldsymbol{y} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. A vector normal to the line is given by $\boldsymbol{n} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$. Let $O = (0, 0)$ denote the origin. The distance from the point to the line is the absolute value of the component of the vector $\boldsymbol{OP}$ in the normal direction

$$d = |\operatorname{comp}_{\boldsymbol{n}} \boldsymbol{OP}| = \left| \frac{\boldsymbol{n} \cdot \boldsymbol{OP}}{\|\boldsymbol{n}\|} \right| = \left| \frac{10}{\sqrt{20}} \right| = \sqrt{5}$$

Exercise 9.11

Find the distance from the point $P = (-2, 3) \in \mathbb{R}^2$ and the line that contains the points $A = (-3, -1)$ and $B = (1, 2)$.

Let V be an inner product space. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in V where $\boldsymbol{y} \neq \boldsymbol{0}$. Let $c = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle}$. Since

$$\langle \boldsymbol{x} - c\boldsymbol{y}, c\boldsymbol{y} \rangle = c \langle \boldsymbol{x} - c\boldsymbol{y}, \boldsymbol{y} \rangle = 0$$

we have by the Pythagorean theorem that

$$\|\boldsymbol{x}\|^2 = \|\boldsymbol{x} - c\boldsymbol{y}\|^2 + \|c\boldsymbol{y}\|^2 \geq \|c\boldsymbol{y}\|^2$$

Using exercise 9.5 and plugging in $c = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle}$ gives

$$\|\boldsymbol{x}\|^2 \geq \|c\boldsymbol{y}\|^2 = |c|^2 \|\boldsymbol{y}\|^2 = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2}{\|\boldsymbol{y}\|^4} \|\boldsymbol{y}\|^2$$

After taking square roots we have justified the following result (the case where $\boldsymbol{y} = \boldsymbol{0}$ is easy).

Lemma 9.2.1: Cauchy-Schwarz Inequality

Let V be an inner product space. Then for all $\boldsymbol{x}, \boldsymbol{y} \in V$ we have that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\| \tag{9.9}$$

Exercise 9.12

Let V be an inner product space. Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in V$ we have that

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$$

if and only if $\boldsymbol{x} = c\boldsymbol{y}$ or $\boldsymbol{y} = c\boldsymbol{x}$.

Exercise 9.13

Let $V = \mathbb{R}^n$ with the standard inner product. Let $\boldsymbol{x} \in \mathbb{R}^n$. Use the Cauchy-Schwarz inequality to prove

$$|x_1| + |x_2| + \cdots + |x_n| \leq \sqrt{n}\,\|\boldsymbol{x}\|$$

Let V be an inner product space and let $\boldsymbol{x}, \boldsymbol{y} \in V$. Using the Cauchy-Schwarz inequality we note

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y} \rangle \\
&= \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \|\boldsymbol{y}\|^2 \\
&\leq \|\boldsymbol{x}\|^2 + 2\langle|\boldsymbol{x}, \boldsymbol{y}\rangle| + \|\boldsymbol{y}\|^2 \\
&\leq \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \\
&= (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2
\end{aligned}$$

After taking square roots we have justified the following result.

Lemma 9.2.2: Triangle Inequality

Let V be an inner product space. For all $\boldsymbol{x}, \boldsymbol{y} \in V$ we have

$$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$

9.3. Orthogonal and Orthonormal Bases

Let V be an inner product space and let $\boldsymbol{u} \in V$. If $\|\boldsymbol{u}\| = 1$, we call $\boldsymbol{u}$ a *unit vector*. If $\boldsymbol{v} \in V$ and $\boldsymbol{v} \neq \boldsymbol{0}$, then $\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ is a unit vector since by exercise 9.5 we have

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

Definition 9.8

Let V be an inner product space. A subset S of V

(1) is *orthogonal* if any two vectors in S are orthogonal.
(2) is an *orthogonal basis* for V if it is orthogonal and a basis for V.
(3) is *orthonormal* if S is orthogonal and consists entirely of unit vectors.
(4) is an *orthonormal basis* for V if it is orthonormal and a basis for V.

The set $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

Let V be an inner product space and $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ an orthogonal set of n *nonzero* vectors in V. Suppose that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = \boldsymbol{0}$$

For any i in $1, \dots, n$ we have

$$\begin{aligned} 0 &= \langle \mathbf{0}, \boldsymbol{v}_i \rangle \\ &= \langle c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n, \boldsymbol{v}_i \rangle \\ &= \langle c_1\boldsymbol{v}_1, \boldsymbol{v}_i \rangle + \cdots \langle c_{i-1}\boldsymbol{v}_{i-1}, \boldsymbol{v}_i \rangle + \langle c_i\boldsymbol{v}_i, \boldsymbol{v}_i \rangle + \langle c_{i+1}\boldsymbol{v}_{i+1}, \boldsymbol{v}_i \rangle + \cdots \langle c_1\boldsymbol{v}_n, \boldsymbol{v}_i \rangle \\ &= \langle c_i\boldsymbol{v}_i, \boldsymbol{v}_i \rangle \\ &= c_i \|\boldsymbol{v}_i\|^2 \end{aligned}$$

and hence $c_i = 0$ since $\|\boldsymbol{v}_i\| \neq 0$. We have justified the following result.

Lemma 9.3.1

Let V be an inner product space and $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ an orthogonal set of n *nonzero* vectors in V. Then S is a linearly independent set of n vectors in V.

Consider the set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_4 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

It is easy to check that S is orthogonal and hence an orthogonal basis for $\mathbb{R}^4$ by lemma 9.3.1 and corollary 5.3.2. Similarily, the set $T = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\}$ where

$$\boldsymbol{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \qquad \boldsymbol{u}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \qquad \boldsymbol{u}_4 = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

is an orthonormal basis for $\mathbb{R}^4$ by by lemma 9.3.1 and corollary 5.3.2.

Exercise 9.14

Let $V = \mathbb{R}^3$ with the standard inner product. Prove that $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

is an orthogonal basis for the subspace

$$W = \{\boldsymbol{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$$

Exercise 9.15

Let $V = C([-1, 1])$ with the inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} f(t)g(t)\, dt$. Prove that $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ where

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}, \qquad \boldsymbol{u}_2 = \sqrt{\frac{3}{2}}t, \qquad \boldsymbol{u}_3 = \sqrt{\frac{5}{8}}(3t^2 - 1)$$

is an orthonormal basis for the subspace $W = P_2(\mathbb{R})$.

Let V be an inner product space and suppose that $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ is an orthonormal basis of n vectors for V. Then for any $\boldsymbol{v} \in V$ we have

$$\boldsymbol{v} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n \tag{9.10}$$

For any i in $1, \ldots, n$ we have

$$\begin{aligned}
\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle &= \langle c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n, \boldsymbol{u}_i \rangle \\
&= \langle c_1\boldsymbol{u}_1, \boldsymbol{u}_i \rangle + \cdots \langle c_{i-1}\boldsymbol{u}_{i-1}, \boldsymbol{u}_i \rangle + \langle c_i\boldsymbol{u}_i, \boldsymbol{u}_i \rangle + \langle c_{i+1}\boldsymbol{u}_{i+1}, \boldsymbol{u}_i \rangle + \cdots \langle c_1\boldsymbol{u}_n, \boldsymbol{u}_i \rangle \\
&= \langle c_i\boldsymbol{u}_i, \boldsymbol{u}_i \rangle \\
&= c_i \|\boldsymbol{u}_i\|^2 \\
&= c_i
\end{aligned}$$

We have justified the following result.

Lemma 9.3.2

Let V be an inner product space and suppose that $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ is an orthonormal basis of n vectors for V. Then for any $\boldsymbol{v} \in V$ we have

$$\boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{u}_1 \rangle \boldsymbol{u}_1 + \langle \boldsymbol{v}, \boldsymbol{u}_2 \rangle \boldsymbol{u}_2 + \cdots + \langle \boldsymbol{v}, \boldsymbol{u}_n \rangle \boldsymbol{u}_n \tag{9.11}$$

Consider again the orthonormal basis $T = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\}$ where

$$\boldsymbol{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \qquad \boldsymbol{u}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \qquad \boldsymbol{u}_4 = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

and let $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. By lemma 9.3.2 we have that

$$\begin{aligned}
\boldsymbol{v} &= \langle \boldsymbol{v}, \boldsymbol{u}_1 \rangle \boldsymbol{u}_1 + \langle \boldsymbol{v}, \boldsymbol{u}_2 \rangle \boldsymbol{u}_2 + \langle \boldsymbol{v}, \boldsymbol{u}_3 \rangle \boldsymbol{u}_3 + \langle \boldsymbol{v}, \boldsymbol{u}_4 \rangle \boldsymbol{u}_4 \\
&= 5\boldsymbol{u}_1 - 2\boldsymbol{u}_2 + \boldsymbol{u}_4
\end{aligned}$$

Exercise 9.16

Let V be an inner product space and suppose that $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is an orthogonal basis of n vectors for V. Prove that for any vector $\boldsymbol{v} \in V$ we have that

$$\boldsymbol{v} = \frac{\langle \boldsymbol{v}, \boldsymbol{v}_1 \rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 + \frac{\langle \boldsymbol{v}, \boldsymbol{v}_2 \rangle}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 + \cdots + \frac{\langle \boldsymbol{v}, \boldsymbol{v}_n \rangle}{\|\boldsymbol{v}_n\|^2}\boldsymbol{v}_n \tag{9.12}$$

Exercise 9.17

Let $V = \mathbb{R}^3$ with the standard inner product.

(1) Prove that $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ -1/2 \\ 1/2 \end{bmatrix}$$

(2) Let $\boldsymbol{v} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$. Use equation (9.12) to calculate c_1, c_2, c_3 such that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3$$

(3) Verify that your coefficients c_1, c_2, c_3 work.

9.4. Orthogonal Complements

Definition 9.9

Let V be an inner product space and S be a subset of V. We define the *orthogonal complement* of S as

$$S^\perp = \{\boldsymbol{x} \in V : \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \text{ for all } \boldsymbol{y} \in S\}$$

Let $V = \mathbb{R}^3$ with the standard inner product and let $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. Then by definition we have

$$S^\perp = \{\boldsymbol{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$$

Thus $S^\perp$ is a plane through the origin. Since $S^\perp = N(A)$ where $A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$, we have that $S^\perp$ is a subspace of $\mathbb{R}^3$. By the rank-nullity theorem we have

$$\dim(S^\perp) = 3 - \operatorname{rank}(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}) = 3 - 1 = 2$$

and it is easy to check that

$$\beta = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a basis for $S^\perp$. The fact that $S^\perp$ in this example is a subspace of $\mathbb{R}^3$ is not an accident. We leave the proof of the following result as an exercise.

Lemma 9.4.1

Let V be an inner product space and S be a subset of V. Then $S^\perp$ is a subspace of V.

Exercise 9.18

Prove lemma 9.4.1.

Let V be an inner product space and W a subspace of V that is generated by $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Let $\boldsymbol{v} \in V$. If $\boldsymbol{v} \in W^\perp$, then $\langle \boldsymbol{v}, \boldsymbol{v}_i \rangle = 0$ for all i in $1, \ldots, n$ since $S \subseteq W$. Next suppose that $\langle \boldsymbol{v}, \boldsymbol{v}_i \rangle = 0$ for all i in $1, \ldots, n$ and let $\boldsymbol{w} \in W$. Since S generates W we have that

$$\boldsymbol{w} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

for some scalars $c_1, c_2, \ldots, c_n$. But then we have that

$$\begin{aligned} \langle \boldsymbol{v}, \boldsymbol{w} \rangle &= \langle \boldsymbol{v}, c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n \rangle \\ &= \langle \boldsymbol{v}, c_1\boldsymbol{v}_1 \rangle + \langle \boldsymbol{v}, c_2\boldsymbol{v}_2 \rangle + \cdots + \langle \boldsymbol{v}, c_n\boldsymbol{v}_n \rangle \\ &= c_1\langle \boldsymbol{v}, \boldsymbol{v}_1 \rangle + c_2\langle \boldsymbol{v}, \boldsymbol{v}_2 \rangle + \cdots + c_n\langle \boldsymbol{v}, \boldsymbol{v}_n \rangle \\ &= 0 \end{aligned}$$

We have justified the following result.

Lemma 9.4.2

Let V be an inner product space and let W be a subspace of V that is generated by the set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Then for any $\boldsymbol{v} \in V$ we have that $\boldsymbol{v} \in W^\perp$ if and only if $\langle \boldsymbol{v}, \boldsymbol{v}_i \rangle = 0$ for all i in $1, \ldots, n$.

Let $V = \mathbb{R}^4$ with the standard inner product and suppose that $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$. Let $W = \operatorname{span}(\{\boldsymbol{w}_1, \boldsymbol{w}_2\})$. By lemma 9.4.2 we have that $\boldsymbol{x} \in W^\perp$ if and only if $\boldsymbol{x} \cdot \boldsymbol{w}_1 = 0$ and $\boldsymbol{x} \cdot \boldsymbol{w}_2 = 0$. Thus $\boldsymbol{x} \in W^\perp$ if and only if $\boldsymbol{x} \in N(A)$ where

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

By the rank-nullity theorem we have that

$$\dim(W^\perp) = \operatorname{nullity}(A) = 4 - \operatorname{rank}(A) = 4 - 2 = 2$$

Adding row 2 to row 1 gives

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

It is easy to check that

$$\beta = \left\{ \begin{bmatrix} -3 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a linearly independent set of 2 vectors in $W^\perp = N(B)$ and hence β is a basis for $W^\perp$.

Exercise 9.19

Let $V = \mathbb{R}^4$ with the standard inner product and suppose that $W = \operatorname{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\})$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}$$

Find a basis for $W^\perp$.

Exercise 9.20

Let $V = \mathbb{R}^m$ with the standard inner product and let $A \in \mathbb{R}^{m \times n}$. Prove that

$$(\operatorname{col}(A))^\perp = N(A^T) \tag{9.13}$$

Let $V = \mathbb{R}^n$ with the standard inner product and let $A \in \mathbb{R}^{m \times n}$. Plugging $A = A^T$ into equation (9.13) gives the formula

$$(\operatorname{col}(A^T))^\perp = N(A) \tag{9.14}$$

CHAPTER 10

Projections and Least Squares

10.1. Orthogonal Projections

Let $V = \mathbb{R}^3$ with the standard inner product. Let W be the plane through the origin in $\mathbb{R}^3$ given by

$$W = \{\boldsymbol{x} \in \mathbb{R}^3 : x_2 - x_3 = 0\}$$

Consider the problem of finding a vector $\boldsymbol{x} \in W$ that is closest to the vector $\boldsymbol{y} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$. In particular, we are looking for $\boldsymbol{x} \in W$ such that

$$\|\boldsymbol{y} - \boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{v}\| \tag{10.1}$$

for all $\boldsymbol{v} \in W$. Intuition suggests that $\|\boldsymbol{y} - \boldsymbol{x}\|$ will be minimized when $\boldsymbol{y} - \boldsymbol{x}$ is orthogonal to the plane W. Thus, we are looking for a vector $\boldsymbol{x}$ such that

$$\boldsymbol{x} \in W, \qquad \boldsymbol{y} - \boldsymbol{x} \in W^{\perp} \tag{10.2}$$

We start by finding a basis for W. Since $W = N(A)$ where $A = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}$ we have that

$$\dim(W) = \text{nullity}(A) = 3 - \text{rank}(A) = 3 - 1 = 2$$

It is easy to check that $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ where $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is a linearly independent set of 2 vectors in W and hence β is a basis for W. Since β is a basis, the $\boldsymbol{x} \in W$ we are looking for can be written as

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$$

for some constants $c_1, c_2 \in \mathbb{R}$ to be determined. Since β generates W, lemma 9.4.2 says the constraint that $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$ is equivalent to

$$(\boldsymbol{y} - c_1\boldsymbol{v}_1 - c_2\boldsymbol{v}_2) \cdot \boldsymbol{v}_1 = 0, \qquad (\boldsymbol{y} - c_1\boldsymbol{v}_1 - c_2\boldsymbol{v}_2) \cdot \boldsymbol{v}_2 = 0 \tag{10.3}$$

After simplifying, the system of equations (10.3) is equivalent to

$$c_1 = -1, \qquad 2c_2 = 2 \tag{10.4}$$

which has the unique solution $c_1 = -1$ and $c_2 = 1$. Thus, $\boldsymbol{x} = -1\boldsymbol{v}_1 + 1\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ is the unique vector satisfying equation (10.2). Furthermore, our intuition suggests that $\boldsymbol{x}$ is the unique vector in W satisfying equation (10.1) for all $\boldsymbol{v} \in W$.

Let $V = C([-1, 1])$ with inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let $W = P_4(\mathbb{R})$ and consider the problem of finding a polynomial $\boldsymbol{f} \in W$ that is closest to the function $\boldsymbol{g} = e^{t^2}$ in V. In particular, we are looking for $\boldsymbol{f} \in W$ such that

$$\|\boldsymbol{g} - \boldsymbol{f}\| \leq \|\boldsymbol{g} - \boldsymbol{h}\| \tag{10.5}$$

for all polynomials $\boldsymbol{h} \in W$. As in the plane example, it turns out that the unique polynomial $\boldsymbol{f} \in W$ that satisfies equation (10.5) for all $\boldsymbol{h} \in W$ is the unique polynomial $\boldsymbol{f}$ satisfying

$$\boldsymbol{f} \in W, \qquad \boldsymbol{g} - \boldsymbol{f} \in W^{\perp} \tag{10.6}$$

Let V be an inner product space and W a finite-dimensional subspace of V. Let $\boldsymbol{y} \in V$ and consider the problem of finding a vector $\boldsymbol{x}$ such that

$$\boldsymbol{x} \in W, \qquad \boldsymbol{y} - \boldsymbol{x} \in W^{\perp} \tag{10.7}$$

Let $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a basis for W containing n vectors. Any vector $\boldsymbol{x} \in W$ can be written

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

Note that by lemma 9.4.2 we have $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$ if and only if

$$\langle \boldsymbol{y} - c_1\boldsymbol{v}_1 - c_2\boldsymbol{v}_2 - \cdots - c_n\boldsymbol{v}_n, \boldsymbol{v}_i \rangle = 0$$

for all $i \in 1, \ldots, n$ or or equivalently $A\boldsymbol{c} = \boldsymbol{z}$ where

$$A = \begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_n, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_n, \boldsymbol{v}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_n \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_n \rangle & \cdots & \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle \end{bmatrix}, \qquad \boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \qquad \boldsymbol{z} = \begin{bmatrix} \langle \boldsymbol{y}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{y}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{y}, \boldsymbol{v}_n \rangle \end{bmatrix}$$

Next suppose that $A\boldsymbol{c} = \boldsymbol{0}$ for some $\boldsymbol{c} \in \mathbb{R}^n$. Then for all i in $1, \ldots, n$ we have

$$0 = c_1\langle \boldsymbol{v}_1, \boldsymbol{v}_i \rangle + c_2\langle \boldsymbol{v}_2, \boldsymbol{v}_i \rangle + \cdots + c_n\langle \boldsymbol{v}_n, \boldsymbol{v}_i \rangle = \langle c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n, \boldsymbol{v}_i \rangle$$

But then

$$\langle c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n, c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n \rangle = 0$$

and hence

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = 0$$

Since β is linearly independent, we have $\boldsymbol{c} = \boldsymbol{0}$. Finally, we note that since $N(A) = \{\boldsymbol{0}\}$ we have that A is invertible by exercise 5.19 and hence

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

where $\boldsymbol{c} = A^{-1}\boldsymbol{z}$ is the unique vector satisfying equation (10.7). We have justified the following theorem.

Theorem 10.1

Let V be an inner product space and W a finite-dimensional subspace of V. For all $\boldsymbol{y} \in V$ there exists a *unique* vector $\boldsymbol{x} \in W$ such that $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$.

Definition 10.1

Let V be an inner product space, W a finite-dimensional subspace of V, and $\boldsymbol{y}$ a vector in V. Let $\boldsymbol{x}$ be the unique vector in W such that $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$. The vector $\boldsymbol{x}$ is called the *orthogonal projection* of $\boldsymbol{y}$ onto W and is denoted $\operatorname{proj}_W \boldsymbol{y}$.

Let V be an inner product space and W a finite-dimensional subspace of V. Since any vector $\boldsymbol{y} \in V$ can be written

$$\boldsymbol{y} = \operatorname{proj}_W \boldsymbol{y} + (\boldsymbol{y} - \operatorname{proj}_W \boldsymbol{y})$$

we have that

$$V = W + W^{\perp}$$

Suppose $\boldsymbol{v} \in W \cap W^{\perp}$. Then $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0$ and hence $W \cap W^{\perp} = \{\boldsymbol{0}\}$. By exercise 3.10 we have justified the following result.

Corollary 10.1.1

If V is an inner product space and W is a finite-dimensional subspace of V, then

$$V = W \oplus W^{\perp}$$

Combining corollary 10.1.1 with exercise 5.8 gives the following result.

Corollary 10.1.2

If V is a finite-dimensional inner product space and W is a subspace of V, then

$$\dim(V) = \dim(W) + \dim(W^{\perp})$$

The next exercise gives an alternate proof of corollary 10.1.2.

Exercise 10.1

Let V be an inner product space and W a finite-dimensional subspace of V. We define the map $T : V \to V$ for $\boldsymbol{y} \in V$ by

$$T\boldsymbol{y} = \text{proj}_W \boldsymbol{y}$$

The map is well defined since the projection of $\boldsymbol{y}$ onto W is unique.

(1) Prove that T is a linear map.
(2) Prove that $R(T) = W$ and $N(T) = W^{\perp}$.
(3) Conclude using the rank-nullity theorem that if V is finite-dimensional then

$$\dim(V) = \dim(W) + \dim(W^{\perp})$$

Exercise 10.2

Let V be a finite-dimensional inner product space and let W be a subspace of V. Prove that

$$W = (W^{\perp})^{\perp}$$

Let $V = \mathbb{R}^m$ with the standard inner product. Let $A \in \mathbb{R}^{m\times n}$. Then $W = \text{col}(A)$ is a subspace of V. By corollary 10.1.2 and exercise 9.20 we have that

$$m = \text{rank}(A) + \text{nullity}(A^T) \tag{10.8}$$

The rank-nullity theorem applied to A^T gives

$$m = \text{rank}(A^T) + \text{nullity}(A^T) \tag{10.9}$$

Combining equations (10.8) and (10.9) gives an alternate proof of theorem 5.4 which says that

$$\text{rank}(A) = \text{rank}(A^T)$$

Let V be an inner product space and W a finite-dimensional subspace of V. In the proof of theorem 10.1 we showed that the the orthogonal projection of a vector $\boldsymbol{y}$ onto W could be calculated by solving a linear system. We will next show that if an orthonormal or othogonal basis is available for W then the orthogonal projection can be calculated directly. Suppose that

$$\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$$

is an orthonormal basis of n vectors for W and let $\boldsymbol{y} \in V$. To find the orthogonal projection of $\boldsymbol{y}$ onto W we look for a vector $\boldsymbol{x} \in W$ such that $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$. Since $\boldsymbol{x} \in W$ we have

$$\boldsymbol{x} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n$$

for some constants $c_1, c_2, \dots, c_n$ to be determined. By lemma 9.4.2 we have $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$ if and only if

$$\langle \boldsymbol{y} - c_1\boldsymbol{u}_1 - c_2\boldsymbol{u}_2 - \cdots - c_n\boldsymbol{u}_n, \boldsymbol{u}_i\rangle = 0$$

for all $i \in 1, \dots, n$ or since β is an orthonormal set

$$c_i = \langle \boldsymbol{y}, \boldsymbol{u}_i\rangle$$

for all $i \in 1, \dots, n$. We have justified the following result.

Theorem 10.2

Let V be an inner product space, W a finite-dimensional subspace of V, and

$$\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$$

an orthonormal basis of n vectors for W. Then for all $\boldsymbol{y} \in V$ we have

$$\operatorname{proj}_W \boldsymbol{y} = \langle \boldsymbol{y}, \boldsymbol{u}_1\rangle \boldsymbol{u}_1 + \langle \boldsymbol{y}, \boldsymbol{u}_2\rangle \boldsymbol{u}_2 + \cdots + \langle \boldsymbol{y}, \boldsymbol{u}_n\rangle \boldsymbol{u}_n$$

The next result gives an easy way to calculate an orthogonal projection using an orthogonal basis for W. We leave the proof as an exercise.

Corollary 10.1.3

Let V be an inner product space, W a finite-dimensional subspace of V, and

$$\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$$

an orthogonal basis of n vectors for W. Then for all $\boldsymbol{y} \in V$ we have

$$\operatorname{proj}_W \boldsymbol{y} = \frac{\langle \boldsymbol{y}, \boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 + \frac{\langle \boldsymbol{y}, \boldsymbol{v}_2\rangle}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 + \cdots + \frac{\langle \boldsymbol{y}, \boldsymbol{v}_n\rangle}{\|\boldsymbol{v}_n\|^2}\boldsymbol{v}_n$$

Let $V = \mathbb{R}^4$ with the standard inner product. Consider the matrix $A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}$ and vector $\boldsymbol{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$. It is easy to check that $\beta = \{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ where $\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{a}_2 = \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix}$ is an orthogonal basis for $W = \operatorname{col}(A)$. Using corollary 10.1.3 we calculate

$$\operatorname{proj}_{\operatorname{col}(A)}\boldsymbol{b} = \frac{\boldsymbol{b}\cdot\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|^2}\boldsymbol{a}_1 + \frac{\boldsymbol{b}\cdot\boldsymbol{a}_2}{\|\boldsymbol{a}_2\|^2}\boldsymbol{a}_2 = \frac{8}{4}\boldsymbol{a}_1 + \frac{45}{90}\boldsymbol{a}_2 = \begin{bmatrix} -1 \\ 1 \\ 2.5 \\ 5.5 \end{bmatrix}$$

To check our result we note that

$$(\boldsymbol{b} - \operatorname{proj}_{\operatorname{col}(A)}) \cdot \boldsymbol{a}_1 = \begin{bmatrix} 0 \\ 1 \\ -1.5 \\ 0.5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 0, \qquad (\boldsymbol{b} - \operatorname{proj}_{\operatorname{col}(A)}) \cdot \boldsymbol{a}_2 = \begin{bmatrix} 0 \\ 1 \\ -1.5 \\ 0.5 \end{bmatrix} \cdot \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix} = 0$$

which shows that $\boldsymbol{b} - \operatorname{proj}_{\operatorname{col}(A)}$ in $(\operatorname{col}(A))^{\perp}$ as expected.

Exercise 10.3

Let $V = \mathbb{R}^3$ with the standard inner product and suppose that

$$\boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Let $W = \operatorname{span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\})$.

(1) Prove that $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal basis for W.
(2) Calculate $\boldsymbol{x} = \operatorname{proj}_W \boldsymbol{y}$.
(3) Verify that $\boldsymbol{y} - \boldsymbol{x}$ is in $W^{\perp}$.

10.2. The Gram-Schmidt Process

Let V be an inner product space and suppose that $S = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n\}$ is a linearly independent set of n vectors in V. For k in $1, \ldots, n$ define

$$S_k = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} \tag{10.10}$$

and

$$T_k = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \tag{10.11}$$

where $\boldsymbol{v}_1 = \boldsymbol{w}_1$ and for k in $2, \ldots, n$

$$\boldsymbol{v}_k = \boldsymbol{w}_k - \mathrm{proj}_{\mathrm{span}(T_{k-1})}\boldsymbol{w}_k \tag{10.12}$$

We will prove by mathematical induction that for all k in $1, \ldots, n$ we have that T_k is an orthogonal set of k nonzero vectors in $\mathrm{span}(S_k)$. The result is clear for $k = 1$. Suppose that the result holds for some k in $1, \ldots, n-1$. We will prove the result holds for $k+1$. We first note that $\boldsymbol{v}_{k+1} \neq \boldsymbol{0}$ since otherwise by equation (10.12)

$$\boldsymbol{w}_{k+1} = \mathrm{proj}_{\mathrm{span}(T_k)}\boldsymbol{w}_{k+1} \in \mathrm{span}(T_k) \subseteq \mathrm{span}(S_k) \tag{10.13}$$

which would contradict the fact that S_{k+1} is a linearly independent set. Combining this result with the fact that $\boldsymbol{v}_{k+1} \in (\mathrm{span}(T_k))^{\perp}$ and the inductive assumption gives us that T_{k+1} is an orthogonal set of $k+1$ nonzero vectors. Finally, we note that by the inductive assumption and equation (10.12) we have that T_{k+1} is in $\mathrm{span}(S_{k+1})$ Combining this result with corollary 10.1.3 proves the following theorem.

Theorem 10.3

Let V be an inner product space and suppose that $S = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n\}$ is a linearly independent subset of n vectors in V. Define $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ where $\boldsymbol{v}_1 = \boldsymbol{w}_1$ and

$$\boldsymbol{v}_k = \boldsymbol{w}_k - \sum_{j=1}^{k-1} \frac{\langle \boldsymbol{w}_k, \boldsymbol{v}_j \rangle}{\|\boldsymbol{v}_j\|^2}\boldsymbol{v}_j, \qquad 2 \leq k \leq n \tag{10.14}$$

Then T is an orthogonal set of n nonzero vectors in $\mathrm{span}(S)$.

The construction of $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ using equation (10.14) is called the *Gram-Schmidt process*.

Let $V = \mathbb{R}^3$ with the standard inner product. Consider the problem of finding an orthogonal basis $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ for $\mathbb{R}^3$ that contains the vector $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. Consider the set $S = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$ where $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and $\boldsymbol{w}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. We note that S is a set of 3 linearly independent vectors and hence is a basis for $\mathbb{R}^3$ containing $\boldsymbol{w}_1$ since

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

The Gram-Schmidt process gives

$$\boldsymbol{v}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{v}_2 = \boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5/6 \\ -1/3 \\ -1/6 \end{bmatrix}$$

$$\boldsymbol{v}_3 = \boldsymbol{w}_3 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix} - \frac{2}{6}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{2}{5}\begin{bmatrix}5/6\\-1/3\\-1/6\end{bmatrix} = \begin{bmatrix}0\\1/5\\-2/5\end{bmatrix}$$

and thus the set $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$ containing $\boldsymbol{w}_1$.

Let $V = \mathbb{R}^4$ with the standard inner product. Consider the problem of finding an orthogonal basis for the subspace

$$W = \{\boldsymbol{x} \in \mathbb{R}^4 : x_1 - x_2 + x_3 - x_4 = 0\}$$

We first find a basis for W. Note that $W = N(A)$ where $A = \begin{bmatrix}1 & -1 & 1 & -1\end{bmatrix}$ and hence

$$\dim(W) = 4 - \text{rank}(A) = 4 - 1 = 3$$

It is easy to check that $S = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$ is a basis for W where $w_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}$, $w_2 = \begin{bmatrix}-1\\0\\1\\0\end{bmatrix}$, and $w_3 = \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$.

To find an orthogonal basis $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ for W we apply the Gram-Schmidt process to the vectors in S. We have

$$\boldsymbol{v}_1 = \boldsymbol{w}_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}$$

$$\boldsymbol{v}_2 = \boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 = \begin{bmatrix}-1\\0\\1\\0\end{bmatrix} - \left(\frac{-1}{2}\right)\begin{bmatrix}1\\1\\0\\0\end{bmatrix} = \begin{bmatrix}-1/2\\1/2\\1\\0\end{bmatrix}$$

$$\boldsymbol{v}_3 = \boldsymbol{w}_3 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\\0\end{bmatrix} - \frac{(-0.5)}{1.5}\begin{bmatrix}-0.5\\0.5\\1\\0\end{bmatrix} = \begin{bmatrix}1/3\\-1/3\\1/3\\1\end{bmatrix}$$

Since β is a linearly independent set of $\dim(W)$ vectors in W, the set β is an orthogonal basis for W.

Exercise 10.4

Let $V = \mathbb{R}^3$ with the standard inner product. Find an orthogonal basis $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ for $\mathbb{R}^3$ that contains the vector $\boldsymbol{w}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$.

Exercise 10.5

Let $V = \mathbb{R}^4$ with the standard inner product. Find an orthogonal basis for the subspace

$$W = \{\boldsymbol{x} \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0\}$$

Exercise 10.6

Let $V = C([-1, 1])$ with the inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Apply the Gram-Schmidt process to the standard basis $S = \{1, t, t^2\}$ for $W = P_2(\mathbb{R})$ to obtain an orthogonal basis T for W. Finally, normalize the vectors in T to obtain an orthonormal basis β for W.

Exercise 10.7

Prove that any finite-dimensional inner product space V has an orthonormal basis.

10.3. Function Approximation

Let V be an inner product space and W be a finite-dimensional subspace of V. Let $\boldsymbol{y} \in V$ and let $\boldsymbol{x} = \operatorname{proj}_W \boldsymbol{y}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Let $\boldsymbol{v}$ be any vector in W. First note that

$$\boldsymbol{y} - \boldsymbol{v} = (\boldsymbol{y} - \boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{v}) \tag{10.15}$$

Since $\boldsymbol{x}, \boldsymbol{v} \in W$ we have $\boldsymbol{x} - \boldsymbol{v} \in W$. Furthermore, since $\boldsymbol{y} - \boldsymbol{x} \in W^{\perp}$ we have that $\langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{v} \rangle = 0$. The Pythagorean theorem gives

$$\|\boldsymbol{y} - \boldsymbol{v}\|^2 = \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \|\boldsymbol{x} - \boldsymbol{v}\|^2 \geq \|\boldsymbol{y} - \boldsymbol{x}\|^2 \tag{10.16}$$

and taking square roots gives

$$\|\boldsymbol{y} - \boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{v}\| \tag{10.17}$$

Next suppose there exists a potentially different vector $\boldsymbol{z} \in W$ such that

$$\|\boldsymbol{y} - \boldsymbol{z}\| \leq \|\boldsymbol{y} - \boldsymbol{v}\| \tag{10.18}$$

for all $\boldsymbol{v} \in W$. Then plugging $\boldsymbol{v} = \boldsymbol{z}$ into equation (10.17) and $\boldsymbol{v} = \boldsymbol{x}$ into equation (10.18) gives

$$\|\boldsymbol{y} - \boldsymbol{x}\| = \|\boldsymbol{y} - \boldsymbol{z}\| \tag{10.19}$$

Next we plug $\boldsymbol{v} = \boldsymbol{z}$ into equation (10.16) to get

$$\|\boldsymbol{y} - \boldsymbol{z}\|^2 = \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \|\boldsymbol{x} - \boldsymbol{z}\|^2 \tag{10.20}$$

Combining equations (10.19) and (10.20) gives $\|\boldsymbol{x} - \boldsymbol{z}\|^2 = 0$ which implies that $\boldsymbol{z} = \boldsymbol{x}$. We have justified the following result.

Theorem 10.4

Let V be an inner product space and W be a finite-dimensional subspace of V. Let $\boldsymbol{y} \in V$ and let $\boldsymbol{x} = \operatorname{proj}_W \boldsymbol{y}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then for all $\boldsymbol{v} \in W$

$$\|\boldsymbol{y} - \boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{v}\|$$

Furthermore, if for some $\boldsymbol{z} \in W$ we have that $\|\boldsymbol{y} - \boldsymbol{z}\| \leq \|\boldsymbol{y} - \boldsymbol{v}\|$ for all $\boldsymbol{v} \in W$, then $\boldsymbol{z} = \boldsymbol{x}$. Thus, $\boldsymbol{x} = \operatorname{proj}_W \boldsymbol{y}$ is the *unique* vector in W that is *closest* to $\boldsymbol{y}$.

Let $V = C([-1, 1])$ with inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let $W = P_2(\mathbb{R})$ and consider the problem of finding a polynomial $\boldsymbol{f} \in W$ that is closest to the function $\boldsymbol{g} = e^{t^2}$ in V. In particular, we are looking for $\boldsymbol{f} \in W$ such that

$$\|\boldsymbol{g} - \boldsymbol{f}\| \leq \|\boldsymbol{g} - \boldsymbol{h}\| \tag{10.21}$$

for all polynomials $\boldsymbol{h} \in W$. By theorem 10.4, the *unique* solution to this minimization problem is

$$\boldsymbol{f} = \operatorname{proj}_{P_2(\mathbb{R})} e^{t^2} \tag{10.22}$$

We will use theorem 10.2 to calculate $\operatorname{proj}_{P_2(\mathbb{R})} e^{t^2}$. To find an orthonormal basis for $P_2(\mathbb{R})$, we run the Gram-Schmidt process on the standard basis $S = \{1, t, t^2\}$ and normalize the resulting vectors. It can be shown (see exericse 10.6) that the resulting orthnormal basis is $\beta = \{u_1, u_2, u_3\}$ where

$$u_1 = \frac{1}{\sqrt{2}}, \qquad u_2 = \sqrt{\frac{3}{2}}t, \qquad u_3 = \sqrt{\frac{5}{8}}(3t^2 - 1)$$

By symmetry we have that

$$\langle e^{t^2}, \boldsymbol{u}_2 \rangle = 0$$

We use mathematical software to estimate the remaining coefficients.

$$\langle e^{t^2}, \boldsymbol{u}_1 \rangle \approx 2.0685, \qquad \langle e^{t^2}, \boldsymbol{u}_3 \rangle \approx 0.66533$$

and set

$$\boldsymbol{f} = \operatorname{proj}_{P_2(\mathbb{R})} e^{t^2} = 2.0685\boldsymbol{u}_1 + 0.66533\boldsymbol{u}_3$$

We use mathematical software to estimate

$$\|e^{t^2} - \boldsymbol{f}\|^2 \approx 0.0075399$$

For comparison, the Maclaurin polynomial approximation in $P_2(\mathbb{R})$ to e^{t^2} is $\boldsymbol{h} = 1 + t^2$. Again using mathematical software we estimate

$$\|e^{t^2} - \boldsymbol{h}\|^2 \approx 0.10037$$

We see that with respect to the given norm, the projection $\boldsymbol{f} = \text{proj}_{P_2(\mathbb{R})} e^{t^2}$ is much closer to $\boldsymbol{g} = e^{t^2}$ then the Maclaurin polynomial approximation $\boldsymbol{h} = 1 + t^2$.

Exercise 10.8

Let $V = C([-1,1])$ with the inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} f(t)g(t)\,dt$ and let $W = P_3(\mathbb{R})$.

(1) Let $\boldsymbol{g} = \sin(\pi t)$. Calculate $\boldsymbol{f} = \text{proj}_{P_3(\mathbb{R})} \sin(\pi t)$, the *orthogonal projection* of $\boldsymbol{g}$ onto $W = P_3(\mathbb{R})$.

Hint: The set $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\}$ is an orthonormal basis for $P_3(\mathbb{R})$ where

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}, \qquad \boldsymbol{u}_2 = \sqrt{\frac{3}{2}}\,t, \qquad \boldsymbol{u}_3 = \sqrt{\frac{5}{8}}(3t^2 - 1), \qquad \boldsymbol{u}_4 = \sqrt{\frac{7}{8}}(5t^3 - 3t)$$

(2) Let $\boldsymbol{h} = \pi t - \dfrac{(\pi t)^3}{6}$ in $P_3(\mathbb{R})$ be the Maclaurin polynomial approximation to $\boldsymbol{g} = \sin(\pi t)$. Using your favorite mathematical software package, plot the functions $\boldsymbol{f}$, $\boldsymbol{g}$, and $\boldsymbol{h}$ on the same graph where $t \in [-1,1]$ and compare.

10.4. Linear Least Squares

Let $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. In the case where $A\boldsymbol{x} = \boldsymbol{b}$ does not have an exact solution, we will seek to find a vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $\|\boldsymbol{b} - A\boldsymbol{x}\|$ is as small as possible.

Definition 10.2

Let $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Then $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is a *least squares solution* to $A\boldsymbol{x} = \boldsymbol{b}$ if

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$$

for all $\boldsymbol{x} \in \mathbb{R}^n$.

Note that any exact solution to $A\boldsymbol{x} = \boldsymbol{b}$ is a least squares solution since if $A\boldsymbol{x} = \boldsymbol{b}$ then $\|\boldsymbol{b} - A\boldsymbol{x}\| = 0$.

We first tackle the question of existence of a least squares solution. Let $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. If $\boldsymbol{b}$ is not in $\text{col}(A)$, then $A\boldsymbol{x} = \boldsymbol{b}$ does not have an exact solution. In this case we will convert the linear system to

$$A\boldsymbol{x} = \text{proj}_{\text{col}(A)}\boldsymbol{b} \tag{10.23}$$

which *does* have an exact solution. Let $\hat{\boldsymbol{x}}$ be any solution to the linear system (10.23) and let $\boldsymbol{x}$ be any vector in $\mathbb{R}^n$. Since $A\boldsymbol{x} \in \text{col}(A)$ we have by theorem 10.4 that

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| = \|\boldsymbol{b} - \text{proj}_{\text{col}(A)}\boldsymbol{b}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$$

We have justified the following result.

Theorem 10.5

Let $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. For all $\boldsymbol{b} \in \mathbb{R}^m$ there exists a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$.

Next we tackle the question of how to find a least squares solution. Let $A \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ be any least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$. Then we have that

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\| \tag{10.24}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. Since any vector in $\text{col}(A)$ can be written as $A\boldsymbol{x}$ for some $\boldsymbol{x} \in \mathbb{R}^n$ we have that

$$A\hat{\boldsymbol{x}} = \text{proj}_{\text{col}(A)}\boldsymbol{b} \tag{10.25}$$

since the projection is the *unique* vector closest to $\boldsymbol{b}$ by theorem 10.4. By definition of projection we have

$$\boldsymbol{b} - A\hat{\boldsymbol{x}} \in (\text{col}(A))^{\perp} \tag{10.26}$$

and since $(\text{col}(A))^{\perp} = N(A^T)$ by exercise 9.20 we have

$$A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \tag{10.27}$$

Next suppose that $\hat{\boldsymbol{x}}$ is any solution to equation (10.27). then $\boldsymbol{b} - A\hat{\boldsymbol{x}} \in N(A^T)$ and hence equation (10.26) holds. By the uniqueness part of theorem 10.1 equation (10.25) holds which implies equation (10.24) by theorem 10.4. Hence $\hat{\boldsymbol{x}}$ is a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$. We have justified the next result.

Theorem 10.6

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. The vector $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\hat{\boldsymbol{x}}$ is a solution to the *normal equations*

$$A^T A\,\boldsymbol{x} = A^T\,\boldsymbol{b} \tag{10.28}$$

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. By theorem 10.5 there exists a least squares solution $\hat{\boldsymbol{x}}$ to $A\boldsymbol{x} = \boldsymbol{b}$ and by theorem 10.6 we have that $\hat{\boldsymbol{x}}$ is a solution to the normal equations. We have justified the following.

Corollary 10.4.1

Let $A \in \mathbb{R}^{m\times n}$. For all $\boldsymbol{b} \in \mathbb{R}^m$ there exists a solution to the normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$.

For an alternate proof of corollary 10.4.1, see exercise 6.38.

Suppose that

$$A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

It is easy to check that $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent. To find the least-squares solution, we solve

$$A^T A x = A^T b$$

We calculate

$$A^T A = \begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix}, \qquad A^T b = \begin{bmatrix} -4 \\ 11 \end{bmatrix}$$

The least-squares solution is thus

$$\hat{x} = (A^T A)^{-1} A^T b = \frac{1}{11}\begin{bmatrix} 22 & 11 \\ 11 & 6 \end{bmatrix}\begin{bmatrix} -4 \\ 11 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

To calculate $\text{proj}_{\text{col}(A)}\boldsymbol{b}$ we note that

$$\text{proj}_{\text{col}(A)} b = A\hat{x} = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

Exercise 10.9

Find a least squares solution to the inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Also calculate $\text{proj}_{\text{col}(A)}\boldsymbol{b}$.

Let $A \in \mathbb{R}^{m\times n}$. Then $A\boldsymbol{x} = \text{proj}_{\text{col}(A)}\boldsymbol{b}$ has a unique solution for all $\boldsymbol{b} \in \mathbb{R}^m$ if and only if nullity$(A) = 0$ or equivalently rank$(A) = n$. Thus $Ax = \boldsymbol{b}$ has a *unique* least squares solution for all $\boldsymbol{b} \in \mathbb{R}^m$ if and only if rank$(A) = n$. Combining this result with exercises 6.37 and 6.39 gives the following theorem.

Theorem 10.7

Let $A \in \mathbb{R}^{m\times n}$. Then the following statements are equivalent.

(1) $A\boldsymbol{x} = \boldsymbol{b}$ has a *unique* least squares solution for all $\boldsymbol{b} \in \mathbb{R}^m$.
(2) rank$(A) = n$.
(3) rank$(A^TA) = n$.
(4) A^TA is invertible.
(5) The normal equations $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ have a unique solution for all $\boldsymbol{b} \in \mathbb{R}^m$

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be n given data points in $\mathbb{R}^2$ and suppose that $x_i \neq x_j$ for $i \neq j$. Let

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Since the assumptions give rank$(X) = 2$, the linear system $X\boldsymbol{\beta} = \boldsymbol{y}$ has a unique least squares solution by theorem 10.7. Let $\hat{\boldsymbol{\beta}}$ be the unique least squares solution to $X\boldsymbol{\beta} = \boldsymbol{y}$. We define the *least-squares line* to be the line $y = \hat{\beta}_1 + \hat{\beta}_2 x$. Note that

$$\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\| \leq \|\boldsymbol{y} - X\boldsymbol{\beta}\| \tag{10.29}$$

for all $\boldsymbol{\beta} \in \mathbb{R}^2$. Since equation (10.29) is equivalent to $\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 \leq \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2$, equation (10.29) says

$$\sum_{i=1}^{n}(y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i))^2 \leq \sum_{i=1}^{n}(y_i - (\beta_1 + \beta_2 x_i))^2 \tag{10.30}$$

for all $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \mathbb{R}^2$. Thus the least-squares line minimizes the *sum of squares* of the differences between the actual and predicted y-coordinates.

Consider the problem of finding the equation $y = \hat{\beta}_1 + \hat{\beta}_2 x$ of the least-squares line that best fits the data points $(0, 1)$, $(1, 1)$, $(2, 2)$, and $(3, 2)$. To find the least-squares line, we need to find the least-squares solution to the following system:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

We calculate

$$X^TX = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, \qquad X^Ty = \begin{bmatrix} 6 \\ 11 \end{bmatrix}$$

The least-squares solution is thus

$$\hat{\beta} = (X^TX)^{-1}X^Ty = \frac{1}{20}\begin{bmatrix} 14 & -6 \\ -6 & 4 \end{bmatrix}\begin{bmatrix} 6 \\ 11 \end{bmatrix} = \begin{bmatrix} 18/20 \\ 8/20 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.4 \end{bmatrix}$$

Thus, the least-squares line is given by $y = \hat{\beta}_0 + \hat{\beta}_1 x = 0.9 + 0.4x$.

Exercise 10.10

Find the equation $y = \hat{\beta}_1 + \hat{\beta}_2 x$ of the least-squares line that best fits the data points (2,1), (5,2), (7,3), and (8,3). Using your favorite mathematical software package, plot the data points and your least-squares line on the same graph where $x \in [0, 3]$ and compare.

CHAPTER 11

Orthogonal Diagonalization and the Spectral Theorem

11.1. Projection Matrices and Orthogonal Matrices

If $U \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ then

$$U^TU = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_n^T \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1\cdot\boldsymbol{u}_1 & \boldsymbol{u}_1\cdot\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_1\cdot\boldsymbol{u}_n \\ \boldsymbol{u}_2\cdot\boldsymbol{u}_1 & \boldsymbol{u}_2\cdot\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_2\cdot\boldsymbol{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_n\cdot\boldsymbol{u}_1 & \boldsymbol{u}_n\cdot\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n\cdot\boldsymbol{u}_n \end{bmatrix} \tag{11.1}$$

Thus by equation 11.1 we have that $U^TU = I_n$ if and only if $\boldsymbol{u}_i \cdot \boldsymbol{u}_i = 1$ for all i in $1, \dots, n$ and $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0$ for $i \neq j$. We have justified the following result.

Lemma 11.1.1

Let $U \in \mathbb{R}^{m\times n}$ have column vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$. Then $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ is an orthonormal set if and only if

$$U^TU = I_n$$

Exercise 11.1

Let $U \in \mathbb{R}^{m\times n}$. Prove that if $U^TU = I_n$, then $n \leq m$.

Exercise 11.2

Let $U \in \mathbb{R}^{m\times n}$ and suppose that $U^TU = I_n$. Prove that

(1) $U\boldsymbol{x} \cdot U\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$ for every $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$
(2) $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for every $\boldsymbol{x}$ in $\mathbb{R}^n$

Suppose $U \in \mathbb{R}^{m\times n}$ has column vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ and $U^TU = I_n$. By lemmas 11.1.1 and 9.3.1, the set of column vectors $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ is an orthonormal basis for $\operatorname{col}(U)$. Thus, for any $\boldsymbol{b} \in \mathbb{R}^m$ we have by equation (1.15) and theorem 10.2 that

$$\begin{aligned} UU^T\boldsymbol{b} &= (\boldsymbol{u}_1\boldsymbol{u}_1^T + \boldsymbol{u}_2\boldsymbol{u}_2^T + \cdots + \boldsymbol{u}_n\boldsymbol{u}_n^T)\boldsymbol{b} \\ &= \boldsymbol{u}_1\boldsymbol{u}_1^T\boldsymbol{b} + \boldsymbol{u}_2\boldsymbol{u}_2^T\boldsymbol{b} + \cdots + \boldsymbol{u}_n\boldsymbol{u}_n^T\boldsymbol{b} \\ &= (\boldsymbol{b}\cdot\boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{b}\cdot\boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{b}\cdot\boldsymbol{u}_n)\boldsymbol{u}_n \\ &= \operatorname{proj}_{\operatorname{col}(U)}\boldsymbol{b} \end{aligned}$$

We have justified the following result.

Lemma 11.1.2

Let $U \in \mathbb{R}^{m\times n}$. If $U^TU = I_n$, then for any $\boldsymbol{b} \in \mathbb{R}^m$ we have

$$UU^T\boldsymbol{b} = \operatorname{proj}_{\operatorname{col}(U)}\boldsymbol{b}$$

Consider the matrix

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}$$

and note that $U^T U = I_2$. We calculate

$$UU^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

By lemma 11.1.2, we have

$$\text{proj}_{\text{col}(U)} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Consider the matrix

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

It is easy to check that $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ where

$$\boldsymbol{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

is an orthonormal basis for $\text{col}(A)$. Thus if $U = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}$ we have that $\text{col}(A) = \text{col}(U)$ and $U^T U = I_2$. By lemma 11.1.2, we have

$$\text{proj}_{\text{col}(A)} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \text{proj}_{\text{col}(U)} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = UU^T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

Exercise 11.3

Let $A \in \mathbb{R}^{m \times n}$. Prove that for all $\boldsymbol{b} \in \mathbb{R}^m$ we have

(1) $\boldsymbol{b} \in \text{col}(A)$ if and only if $\text{proj}_{\text{col}(A)}\boldsymbol{b} = \boldsymbol{b}$.

(2) $\boldsymbol{b} \in N(A^T)$ if and only if $\text{proj}_{\text{col}(A)}\boldsymbol{b} = \boldsymbol{0}$.

Exercise 11.4

Consider the matrix

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

(1) Find a matrix $P \in \mathbb{R}^{3 \times 3}$ such that for all $\boldsymbol{b} \in \mathbb{R}^3$ we have

$$P\boldsymbol{b} = \text{proj}_{\text{col}(A)}\boldsymbol{b}$$

(2) Calculate $\text{proj}_{\text{col}(A)} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Definition 11.1

A matrix U in $\mathbb{R}^{n\times n}$ is called an *orthogonal matrix* if $U^TU = I_n$.

Exercise 11.5

Prove that $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix if and only if U is invertible and $U^{-1} = U^T$.

Exercise 11.6

Example: Suppose that

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

Is U an orthogonal matrix? If so, what is U^{-1}?

Exercise 11.7

Prove that $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix if and only if U^T is an orthogonal matrix.

Exercise 11.8

Prove that if $U_1, U_2 \in \mathbb{R}^{n\times n}$ are orthogonal matrices, then U_1U_2 is an orthogonal matrix.

11.2. Orthogonal Diagonalization

Definition 11.2

A matrix $A \in \mathbb{R}^{n\times n}$ is called *orthogonally diagonalizable* if there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that

$$A = UDU^T$$

or equivalently

$$D = U^TAU$$

Note that if A is orthogonally diagonalizable, then A is diagonalizable. Suppose that A is orthogonally diagonalizable. Then since $A = UDU^T$ for some orthogonal matrix $U \in \mathbb{R}^{n\times n}$ and diagonal matrix $D \in \mathbb{R}^{n\times n}$ we have that

$$A^T = (UDU^T)^T = (U^T)^TD^TU^T = UDU^T = A$$

We have justified the following result.

Lemma 11.2.1

If $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable, then $A^T = A$ and hence A is a symmetric matrix.

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Since A has 2 distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, A is diagonalizable by exercise 8.27. However since $A \neq A^T$, the matrix A is not orthogonally diagonalizable by lemma 11.2.1.

By exercise 8.2, we have that a matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if there exists a basis for $\mathbb{R}^n$ consisting of eigenvectors of A. As the example $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ indicates, the existence of a basis for $\mathbb{R}^n$ consisting of eigenvectors of A is *not* a sufficient condition for orthogonal diagonalizability.

Suppose that $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable. Then $A = UDU^T$ for some orthogonal matrix $U \in \mathbb{R}^{n\times n}$ and diagonal matrix $D \in \mathbb{R}^{n\times n}$. Suppose U has column vectors then $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ Then for all i in $1, \ldots, n$ we have

$$U\boldsymbol{e}_i = \boldsymbol{u}_i, \qquad U^T\boldsymbol{u}_i = \boldsymbol{e}_i$$

$$A\boldsymbol{u}_i = UDU^T\boldsymbol{u}_i = UD\boldsymbol{e}_i = UD_{ii}\boldsymbol{e}_i = D_{ii}\boldsymbol{u}_i$$

and thus $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A. We have justified the first part of the following theorem. We leave the remainder of the proof as an exercise.

Theorem 11.1

A matrix $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable if and only if there exists an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A.

Exercise 11.9

Finish the proof of theorem 11.1.

Consider the matrix $A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 3 \end{bmatrix}$. Since A is symmetric it could be orthogonally diagonalizable. To apply theorem 11.1, we search for an an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A. We first calculate

$$\begin{aligned} \det\left(\begin{bmatrix} 3-\lambda & 0 & 1 \\ 0 & -\lambda & 0 \\ 1 & 0 & 3-\lambda \end{bmatrix}\right) &= (-\lambda)(-1)^{2+2}[(3-\lambda)^2 - 1] \\ &= (-\lambda)(\lambda^2 - 6\lambda + 8) \\ &= (-\lambda)(\lambda - 2)(\lambda - 4) \end{aligned}$$

Thus there are three distinct eigenvalues

$$\lambda_1 = 4, \qquad \lambda_2 = 2, \qquad \lambda_3 = 0$$

To find a unit length eigenvector for $\lambda_1 = 4$, we solve

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & -1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

to get $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. Thus we set $\boldsymbol{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$. To find a unit length eigenvector for $\lambda_2 = 2$, we solve

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

to get $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$. Thus we set $\boldsymbol{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}$. To find a unit length eigenvector for $\lambda_3 = 0$, we solve

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

to get $\boldsymbol{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Thus we set $\boldsymbol{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. It is easy to check that $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A and hence A is orthogonally diagonalizable by theorem 11.1. In fact, setting

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

gives

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

$$\begin{aligned} UDU^T &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 4/\sqrt{2} & 2/\sqrt{2} & 0 \\ 0 & 0 & 0 \\ 4/\sqrt{2} & -2/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 3 \end{bmatrix} \\ &= A \end{aligned}$$

The fact that eigenvectors from different eigenspaces in the previous example are orthogonal is not a coincidence. Let $A \in \mathbb{R}^{n \times n}$ such that $A^T = A$. Suppose that $A\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1$, $A\boldsymbol{u}_2 = \lambda_2 \boldsymbol{u}_2$, and $\lambda_1 \neq \lambda_2$. First we note that

$$\begin{aligned} \lambda_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) &= (\lambda_1 \boldsymbol{u}_1) \cdot \boldsymbol{u}_2 \\ &= (\lambda_1 \boldsymbol{u}_1)^T \boldsymbol{u}_2 \\ &= (A\boldsymbol{u}_1)^T \boldsymbol{u}_2 \\ &= (\boldsymbol{u}_1^T A^T)\boldsymbol{u}_2 \\ &= \boldsymbol{u}_1^T (A^T \boldsymbol{u}_2) \\ &= \boldsymbol{u}_1^T (A\boldsymbol{u}_2) \\ &= \boldsymbol{u}_1^T (\lambda_2 \boldsymbol{u}_2) \\ &= \boldsymbol{u}_1 \cdot (\lambda_2 \boldsymbol{u}_2) \\ &= \lambda_2(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) \end{aligned}$$

Thus we have $\lambda_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) = \lambda_2(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2)$ and hence

$$0 = \lambda_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) - \lambda_2(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) = (\lambda_1 - \lambda_2)(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2)$$

Finally, since $\lambda_1 - \lambda_2 \neq 0$ we must have

$$\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 0$$

We have justifed the following theorem.

Theorem 11.2

Let $A \in \mathbb{R}^{n \times n}$ such that $A^T = A$. Then any two eigenvectors corresponding to distinct eigenvalues of A are orthogonal.

Exercise 11.10

Let $A \in \mathbb{R}^{n\times n}$ such that $A^T = A$ and suppose A has n *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Prove that A is orthogonally diagonalizable.

Exercise 11.11

Suppose that $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$. Is A orthogonally diagonalizable? If so, find $U \in \mathbb{R}^{2\times 2}$ and $D \in \mathbb{R}^{2\times 2}$ such that $U^T U = I_2$ and $A = UDU^T$.

Exercise 11.12

Suppose that $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$. Is A orthogonally diagonalizable? If so, find $U \in \mathbb{R}^{3\times 3}$ and $D \in \mathbb{R}^{3\times 3}$ such that $U^T U = I_3$ and $A = UDU^T$.

As the next example shows, a symmetric matrix $A \in \mathbb{R}^{n\times n}$ can be orthogonally diagonalizable even if A does not have n distinct eigenvalues. Consider the matrix $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$. Using mathematical software we determine that A has two distinct eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 5$. For $\lambda_1 = 2$ we solve

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

It is easy to check that $S = \{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ where

$$\boldsymbol{w}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{w}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

is a basis for E_{λ_1}. Running the Gram-Schmidt process on S gives us an orthogonal basis for E_{λ_1} given by $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

Normalizing the vectors in T gives us an orthonormal basis for E_{λ_1} given by $\beta_1 = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ where

$$\boldsymbol{u}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}$$

For $\lambda_2 = 5$ we solve

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

It is easy to check that $\beta_2 = \{\boldsymbol{u}_3\}$ where $\boldsymbol{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$ is an orthonormal basis for E_{λ_2}. By theorem 11.2,

$$\beta = \beta_1 \cup \beta_2 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$$

is an orthonormal basis for $\mathbb{R}^3$ consisting of eigenvectors of A and hence A is orthogonally diagonalizable by theorem 11.1.

Exercise 11.13

Suppose that

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

Find an orthogonal matrix U and a diagonal matrix D such that

$$A = UDU^T$$

Hint: A has two distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 4$.

11.3. The Spectral Theorem

Let $A \in \mathbb{R}^{n\times n}$. By lemma 11.2.1, a necessary condition for A to be orthogonally diagonalizable is that $A^T = A$. The fact that this necessary condition is also *sufficient* is a consequence of the *spectral theorem* which we prove in this section.

In order to prove the spectral theorem we will need the result that any real symmetric matrix must have a real eigenvalue. This is not an easy result and requires some ideas from analysis. For reference we will give the proof of the following lemma at the end of this chapter.

Lemma 11.3.1: Spectral Lemma

Let $A \in \mathbb{R}^{n\times n}$ and suppose that $A^T = A$. Then there exists $\lambda \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$ such that

$$A\boldsymbol{x} = \lambda\boldsymbol{x}, \qquad \|\boldsymbol{x}\| = 1$$

We will next show that for all k in $1, \ldots, n$ there exists an orthonormal set of k eigenvectors of A using mathematical induction on k. The base case where $k = 1$ is handled by the spectral lemma. Now suppose that the theorem is true for some integer k satisfying $1 \le k < n$. We prove the result for $k+1$. By the inductive assumption, there exists an orthonormal set $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ in $\mathbb{R}^n$ consisting of k eigenvectors of A. Let $T = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ that contains S. Let $U \in \mathbb{R}^{n\times(n-k)}$ be the matrix with column vectors $\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n$. Since

$$(U^T AU)^T = U^T A^T (U^T)^T = U^T AU$$

there exists $\boldsymbol{x} \in \mathbb{R}^{n-k}$ and $\lambda \in \mathbb{R}$ such that

$$U^T AU\boldsymbol{x} = \lambda\boldsymbol{x}, \qquad \|\boldsymbol{x}\| = 1 \tag{11.2}$$

by the spectral lemma. We will show that $S \cup \{U\boldsymbol{x}\}$ is an orthonormal set in $\mathbb{R}^n$ consisting of $k+1$ eigenvectors of A. First note that $U\boldsymbol{x} \in \mathbb{R}^n$ and that

$$\|U\boldsymbol{x}\| = \|\boldsymbol{x}\| = 1$$

by exercise 11.2. Next note that for all i in $1, \ldots, k$ we have

$$U\boldsymbol{x} \cdot \boldsymbol{u}_i = \boldsymbol{x}^T U^T \boldsymbol{u}_i = \boldsymbol{x}^T \boldsymbol{0} = 0$$

since T is an orthonormal set. We have shown that $S \cup \{U\boldsymbol{x}\}$ is an orthonormal set of $k+1$ vectors in $\mathbb{R}^n$ so it remains to show that $U\boldsymbol{x}$ is an eigenvector of A. We next use the orthonormal basis T to write

$$AU\boldsymbol{x} = (AU\boldsymbol{x}\cdot\boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (AU\boldsymbol{x}\cdot\boldsymbol{u}_k)\boldsymbol{u}_k + (AU\boldsymbol{x}\cdot\boldsymbol{u}_{k+1})\boldsymbol{u}_{k+1} + \cdots + (AU\boldsymbol{x}\cdot\boldsymbol{u}_n)\boldsymbol{u}_n$$

and note that for i in $1, \ldots, k$ we have

$$AU\boldsymbol{x}\cdot\boldsymbol{u}_i = \boldsymbol{x}^T U^T A^T \boldsymbol{u}_i = \boldsymbol{x}^T U^T A\boldsymbol{u}_i = \boldsymbol{x}^T U^T \lambda_i \boldsymbol{u}_i = \lambda_i \boldsymbol{x}^T U^T \boldsymbol{u}_i = 0$$

Thus $AU\boldsymbol{x} \in \mathrm{col}(U)$ and by equation (11.2) and lemma 11.1.2 we have

$$\lambda U\boldsymbol{x} = U(\lambda\boldsymbol{x}) = UU^T AU\boldsymbol{x} = \mathrm{proj}_{\mathrm{col}(U)} AU\boldsymbol{x} = AU\boldsymbol{x}$$

and we have justified the following result.

Lemma 11.3.2

Let $A \in \mathbb{R}^{n\times n}$. If $A^T = A$, then for any k satisfying $1 \le k \le n$, there exists an orthonormal set $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ in $\mathbb{R}^n$ containing k eigenvectors of A.

We now have all the tools necessary to prove the spectral theorem which we leave as an exercise.

Theorem 11.3: Spectral Theorem

Let $A \in \mathbb{R}^{n\times n}$. Then A is orthogonally diagonalizable if and only if $A^T = A$.

Exercise 11.14

Prove theorem 11.3.

Exercise 11.15

Let A in $\mathbb{R}^{n\times n}$ such that $A^T = A$. Let $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n-1}, \boldsymbol{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ where the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n-1}$ are all eigenvectors of A. That is, for all i satisfying $1 \le i \le n-1$ we have that

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$$

for some eigenvalue $\lambda_i \in \mathbb{R}$. Prove that $A\boldsymbol{u}_n \cdot \boldsymbol{u}_i = 0$ for all i satisfying $1 \le i \le n-1$. Write $A\boldsymbol{u}_n$ as a linear combination of vectors in β and conclude that $\boldsymbol{u}_n$ is an eigenvector of A.

11.4. Spectral Lemma Proof

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables. We define

$$\frac{\partial f}{\partial x_i}(\boldsymbol{a}) = \frac{d}{dt} f(\boldsymbol{a} + t\boldsymbol{e}_i)\bigg|_{t=0}, \qquad \nabla f(\boldsymbol{a}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\boldsymbol{a}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}$$

Exercise 11.16

Consider the functions of n variables given by

$$f(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x}, \qquad g(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = \boldsymbol{x} \cdot \boldsymbol{x}$$

where $\boldsymbol{a}$ is a constant vector in $\mathbb{R}^n$. Show that

$$\nabla f = \boldsymbol{a}, \qquad \nabla g = 2\boldsymbol{x}, \qquad \nabla x_i = \boldsymbol{e}_i$$

Let $A \in \mathbb{R}^{n\times n}$ and suppose A is symmetric. Let $f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} = A\boldsymbol{x} \cdot \boldsymbol{x}$. The function $f(\boldsymbol{x})$ is called a *quadratic form*. We will show that $\nabla f = 2A\boldsymbol{x}$. Suppose A has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Using the product rule and sum rule for gradients

$$\nabla(f_1 + \cdots + f_n) = \nabla f_1 + \cdots + \nabla f_n, \qquad \nabla(fg) = g\nabla f + f\nabla g$$

and exercise 11.16 gives

$$\begin{aligned}
\nabla A\boldsymbol{x}\cdot\boldsymbol{x} &= \nabla(x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{x}_n)\cdot\boldsymbol{x} \\
&= \nabla(x_1\boldsymbol{a}_1\cdot\boldsymbol{x} + x_2\boldsymbol{a}_2\cdot\boldsymbol{x} + \cdots + x_n\boldsymbol{a}_n\cdot\boldsymbol{x}) \\
&= \nabla x_1(\boldsymbol{a}_1\cdot\boldsymbol{x}) + \nabla x_2(\boldsymbol{a}_2\cdot\boldsymbol{x}) + \cdots + \nabla x_n(\boldsymbol{a}_n\cdot\boldsymbol{x}) \\
&= (\boldsymbol{a}_1\cdot\boldsymbol{x})\boldsymbol{e}_1 + x_1\boldsymbol{a}_1 + (\boldsymbol{a}_2\cdot\boldsymbol{x})\boldsymbol{e}_2 + x_1\boldsymbol{a}_2 + \cdots + (\boldsymbol{a}_n\cdot\boldsymbol{x})\boldsymbol{e}_n + x_1\boldsymbol{a}_n \\
&= A^T\boldsymbol{x} + A\boldsymbol{x} \\
&= 2A\boldsymbol{x}
\end{aligned}$$

We have justified the following result

Lemma 11.4.1

Let $A \in \mathbb{R}^{n\times n}$ and suppose A is symmetric. Let $f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} = A\boldsymbol{x}\cdot\boldsymbol{x}$. Then $\nabla f = 2A\boldsymbol{x}$.

Consider the *Rayleigh Quotient*

$$f(\boldsymbol{x}) = \frac{A\boldsymbol{x}\cdot\boldsymbol{x}}{\boldsymbol{x}\cdot\boldsymbol{x}} \tag{11.3}$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric. Using the quotient rule for gradients $\nabla(f/g) = (g\nabla f - f\nabla g)/g^2$, exercise 11.16, and lemma 11.4.1 we calculate

$$\nabla f = \frac{(\boldsymbol{x}\cdot\boldsymbol{x})(2A\boldsymbol{x}) - (A\boldsymbol{x}\cdot\boldsymbol{x})(2\boldsymbol{x})}{(\boldsymbol{x}\cdot\boldsymbol{x})^2} \tag{11.4}$$

Let S be the closed and bounded set $S = \{\boldsymbol{x}\in\mathbb{R}^n : \boldsymbol{x}\cdot\boldsymbol{x} = 1\}$. By the extreme value theorem, there exists $\boldsymbol{x}_0 \in S$ such that $f(\boldsymbol{x}_0) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x}\in S$. Let $\boldsymbol{y}\in\mathbb{R}^n - \{\mathbf{0}\}$ be arbitrary and set $\boldsymbol{u} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$. Since

$$f(\boldsymbol{x}_0) \geq f(\boldsymbol{u}) = f(\boldsymbol{y})$$

we have that $f(\boldsymbol{x}_0) \geq f(\boldsymbol{y})$ for all $\boldsymbol{y}$ in $\mathbb{R}^n - \{\mathbf{0}\}$ and thus

$$\mathbf{0} = \nabla f(\boldsymbol{x}_0) = 2A\boldsymbol{x}_0 - (A\boldsymbol{x}_0\cdot\boldsymbol{x}_0)2\boldsymbol{x}_0$$

Setting $\lambda = A\boldsymbol{x}_0\cdot\boldsymbol{x}_0 \in \mathbb{R}$ gives

$$A\boldsymbol{x}_0 = \lambda\boldsymbol{x}_0 \tag{11.5}$$

The above proof of the spectral lemma was taken from a book by Frank Jones at Rice. The spectral lemma can also be shown using the Cauchy-Schwarz inequality or Lagrange Multipliers. The other standard proof is to use the fundamental theorem of algebra and show that a symmetric matrix can only have real eigenvalues.

CHAPTER 12

Singular Value Decomposition

12.1. Derivation and Examples

Let $A \in \mathbb{R}^{m\times n}$ and suppose $AA^T \in \mathbb{R}^{m\times m}$ has eigenvalue $\lambda \in \mathbb{R}$ with corresponding unit length eigenvector $\boldsymbol{u} \in \mathbb{R}^m$. Then

$$0 \le \|A^T\boldsymbol{u}\|^2 = A^T\boldsymbol{u}\cdot A^T\boldsymbol{u} = \boldsymbol{u}^T AA^T\boldsymbol{u} = \lambda\boldsymbol{u}\cdot\boldsymbol{u} = \lambda$$

and we have justified the following result.

Lemma 12.1.1

Let $A \in \mathbb{R}^{m\times n}$ and suppose $AA^T \in \mathbb{R}^{m\times m}$ has eigenvalue $\lambda \in \mathbb{R}$ with corresponding unit length eigenvector $\boldsymbol{u} \in \mathbb{R}^m$. Then

$$\lambda = \|A^T\boldsymbol{u}\|^2$$

and hence $\lambda \ge 0$.

Let $A \in \mathbb{R}^{m\times n}$. Since $AA^T \in \mathbb{R}^{m\times m}$ is symmetric, the spectral theorem gives

$$AA^T = UDU^T = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m\end{bmatrix}\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_m\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m\end{bmatrix}^T \tag{12.1}$$

where $U^TU = I_m$ and $AA^T\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i$ for i in $1,\ldots,m$. By lemma 12.1.1, we can assume that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m \ge 0 \tag{12.2}$$

Suppose $\text{rank}(A) = r$. By exercises 6.33 and 6.38, we have

$$\text{rank}(D) = \text{rank}(UDU^T) = \text{rank}(AA^T) = \text{rank}(A) = r$$

and thus

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0, \qquad \lambda_{r+1} = \cdots = \lambda_m = 0 \tag{12.3}$$

Let

$$U_r = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r\end{bmatrix} \tag{12.4}$$

and note that

$$U_r^TU_r = I_r \tag{12.5}$$

For all i in $1,\ldots,r$ we have $\boldsymbol{u}_i = \frac{1}{\lambda_i}AA^T\boldsymbol{u}_i$ and thus $\boldsymbol{u}_i \in \text{col}(A)$. But then $\text{col}(U_r) \subseteq \text{col}(A)$ and since $\text{rank}(U_r) = r = \text{rank}(A)$ we have by theorem 5.7 that

$$\text{col}(U_r) = \text{col}(A) \tag{12.6}$$

Next suppose A has column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and note that by lemma 11.1.2 we have

$$\begin{aligned}U_rU_r^TA &= \begin{bmatrix}U_rU_r^T\boldsymbol{a}_1 & U_rU_r^T\boldsymbol{a}_2 & \cdots & U_rU_r^T\boldsymbol{a}_n\end{bmatrix}\\ &= \begin{bmatrix}\text{proj}_{\text{col}(U_r)}\boldsymbol{a}_1 & \text{proj}_{\text{col}(U_r)}\boldsymbol{a}_2 & \cdots & \text{proj}_{\text{col}(U_r)}\boldsymbol{a}_n\end{bmatrix}\\ &= \begin{bmatrix}\text{proj}_{\text{col}(A)}\boldsymbol{a}_1 & \text{proj}_{\text{col}(A)}\boldsymbol{a}_2 & \cdots & \text{proj}_{\text{col}(A)}\boldsymbol{a}_n\end{bmatrix}\\ &= A\end{aligned}$$

and thus

$$A = U_r U_r^T A \tag{12.7}$$

For all i in $1, \ldots, r$ define $\boldsymbol{v}_i = \frac{A^T \boldsymbol{u}_i}{\sqrt{\lambda_i}}$ and set

$$V_r = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix} \tag{12.8}$$

By lemma 12.1.1 we have

$$\|\boldsymbol{v}_i\|^2 = \frac{\|A^T \boldsymbol{u}_i\|^2}{\lambda_i} = 1$$

and for all i, j in $1, \ldots, r$ such that $i \neq j$ we have

$$\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}} A^T \boldsymbol{u}_i \cdot A^T \boldsymbol{u}_j = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}} \boldsymbol{u}_i^T A A^T \boldsymbol{u}_j = \frac{\lambda_j}{\sqrt{\lambda_i}\sqrt{\lambda_j}} \boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0$$

Thus we have that

$$V_r^T V_r = I_r \tag{12.9}$$

For all i in $1, \ldots, r$ set $\sigma_i = \sqrt{\lambda_i}$ and define

$$\Sigma_r = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \tag{12.10}$$

We call $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ the *singular values* of A. Next note that

$$A^T U_r = \begin{bmatrix} A^T \boldsymbol{u}_1 & A^T \boldsymbol{u}_2 & \cdots & A^T \boldsymbol{u}_r \end{bmatrix} = \begin{bmatrix} \sigma_1 \boldsymbol{v}_1 & \sigma_2 \boldsymbol{v}_2 & \cdots & \sigma_r \boldsymbol{v}_r \end{bmatrix} = V_r \Sigma_r \tag{12.11}$$

Finally we have by equations (12.7) and (12.11) that

$$A = U_r U_r^T A = U_r (A^T U_r)^T = U_r (V_r \Sigma_r)^T = U_r \Sigma_r V_r^T \tag{12.12}$$

We call

$$A = U_r \Sigma_r V_r^T \tag{12.13}$$

a *singular value decomposition* (SVD) for A. We have justified the following result.

Theorem 12.1

For any $A \in \mathbb{R}^{m \times n}$ such that $\operatorname{rank}(A) = r$, there exists $U_r \in \mathbb{R}^{m \times r}$, $V_r \in \mathbb{R}^{n \times r}$, and a diagonal matrix $\Sigma_r \in \mathbb{R}^{r \times r}$ such that

$$A = U_r \Sigma_r V_r^T, \qquad U_r^T U_r = I_r, \qquad V_r^T V_r = I_r, \qquad \Sigma_r = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$. By inspection we have $r = \operatorname{rank}(A) = 2$. We calculate

$$AA^T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

As expected, we have $\operatorname{rank}(AA^T) = 2 = r$. We next note that

$$\det\left(\begin{bmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & 0 \\ 2 & 0 & 2-\lambda \end{bmatrix} \right) = (2-\lambda)(-1)^{2+2}((2-\lambda)^2 - 4) = (2-\lambda)(\lambda)(\lambda - 4)$$

and the eigenvalues of AA^T are

$$\lambda_1 = 4, \qquad \lambda_2 = 2, \qquad \lambda_3 = 0$$

Thus we have that $\sigma_1 = 2$, $\sigma_2 = \sqrt{2}$ and

$$\Sigma_r = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

Next we find U_r. Starting with $\lambda_1 = 4$ we solve

$$\begin{bmatrix} -2 & 0 & 2 \\ 0 & -2 & 0 \\ 2 & 0 & -2 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

to get $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and hence $\boldsymbol{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$. Next we consider $\lambda_2 = 2$ and solve

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

to get $\boldsymbol{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and hence $\boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Thus we have

$$U_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}$$

Next we find V_r. We calculate

$$\boldsymbol{v}_1 = \frac{A^T \boldsymbol{u}_1}{\sigma_1} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \qquad \boldsymbol{v}_2 = \frac{A^T \boldsymbol{u}_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

Thus we have

$$V_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{bmatrix}$$

Finally we verify that

$$U_r \Sigma_r V_r^T = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = A$$

Exercise 12.1

Find a singular value decomposition (SVD) for the following matrix

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

Verify by direct calculation that $A = U_r \Sigma_r V_r^T$.

Exercise 12.2

Let $A \in \mathbb{R}^{m \times n}$ have SVD $A = U_r \Sigma_r V_r^T$. Prove that $\text{col}(V_r) = \text{col}(A^T)$.

Exercise 12.3

Let $A \in \mathbb{R}^{m\times n}$ have SVD $A = U_r\Sigma_r V_r^T$. Prove that for all $\boldsymbol{b} \in \mathbb{R}^m$ and all $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$U_rU_r^T\boldsymbol{b} = \text{proj}_{\text{col}(A)}\boldsymbol{b}, \qquad V_rV_r^T\boldsymbol{x} = \text{proj}_{\text{col}(A^T)}\boldsymbol{x}$$

Consider again the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ with SVD given by

$$A = U_r\Sigma_r V_r^T = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

We calculate

$$U_rU_r^T = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}, \qquad V_rV_r^T = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}$$

and by exercise 12.3 we have

$$\text{proj}_{\text{col}(A)} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = U_rU_r^T \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \qquad \text{proj}_{\text{col}(A^T)} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} = V_rV_r^T \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

We conclude that the vector $\boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ is in col(A) and the vector $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$ is in $(\text{col}(A^T))^\perp = N(A)$.

Exercise 12.4

Let $A \in \mathbb{R}^{3\times 3}$ where $A = U_r\Sigma_r V_r^T$ and

$$U_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}, \qquad \Sigma_r = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad V_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ -1/\sqrt{2} & 0 \end{bmatrix}$$

(1) What is rank(A)?
(2) Calculate $U_rU_r^T$ and $V_rV_r^T$.
(3) Is the vector $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$ in col(A)? Explain.
(4) Is the vector $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ in col(A^T)? Explain.

Exercise 12.5

Let $A \in \mathbb{R}^{m\times n}$ and suppose $r = \text{rank}(A)$. Suppose A^TA has orthogonal diagonalization

$$A^TA = VDV^T = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}^T$$

where $V^TV = I_n$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

(1) How many nonzero eigenvalues does A^TA have? Explain.
(2) Prove that $\lambda_i \geq 0$ for all i in $1, \ldots, n$. Hint: Calculate $\|A\boldsymbol{v}_i\|^2$.
(3) Prove that $\beta = \{A\boldsymbol{v}_1, A\boldsymbol{v}_2, \ldots, A\boldsymbol{v}_r\}$ is an orthogonal basis for $\text{col}(A)$.

Exercise 12.6

Let $A \in \mathbb{R}^{m\times n}$. Recall by exercise 6.37 that $N(A^TA) = N(A)$ and hence $N(AA^T) = N(A^T)$. Let $\lambda \in \mathbb{R}$ be a nonzero scalar. Prove that λ is an eigenvalue of A^TA if and only if λ is an eigenvalue of AA^T. Conclude that A and A^T have the same singular values.

Exercise 12.7

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $\text{rank}(A) = r$. Prove that

$$\text{rank}(U_rU_r^T) = \text{rank}(V_rV_r^T) = r$$

Exercise 12.8

Let $A \in \mathbb{R}^{10\times 15}$. Suppose you know that $\text{rank}(A) = 10$. Calculate $\text{rank}(U_rU_r^T)$ and $\text{nullity}(U_rU_r^T)$. Calculate $\text{rank}(V_rV_r^T)$ and $\text{nullity}(V_rV_r^T)$.

12.2. The Pseudoinverse

Let $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and suppose that $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Let $\boldsymbol{x} \in \mathbb{R}^n$ be any solution to $A\boldsymbol{x} = \boldsymbol{b}$ and set $\boldsymbol{p} = \text{proj}_{\text{col}(A^T)}\boldsymbol{x}$. Since $\boldsymbol{x} - \boldsymbol{p}$ is in $(\text{col}(A^T))^\perp = N(A)$ we have that

$$\mathbf{0} = A(\boldsymbol{x} - \boldsymbol{p}) = A\boldsymbol{x} - A\boldsymbol{p} \tag{12.14}$$

and thus $\boldsymbol{p} \in \text{col}(A^T)$ is a solution to $A\boldsymbol{x} = \boldsymbol{b}$ since

$$A\boldsymbol{p} = A\boldsymbol{x} = \boldsymbol{b} \tag{12.15}$$

Suppose $\boldsymbol{y} \in \mathbb{R}^n$ is a potentially different vector satisfying

$$\boldsymbol{y} \in \text{col}(A^T), \qquad A\boldsymbol{y} = \boldsymbol{b}$$

Then $\boldsymbol{p} - \boldsymbol{y} \in \text{col}(A^T)$ and since

$$A(\boldsymbol{p} - \boldsymbol{y}) = A\boldsymbol{p} - A\boldsymbol{y} = \boldsymbol{b} - \boldsymbol{b} = \mathbf{0}$$

we have that $\boldsymbol{p} - \boldsymbol{y}$ in $N(A) = (\text{col}(A^T))^\perp$. But then $\boldsymbol{p} - \boldsymbol{y}$ is perpendicular to itself and thus $\boldsymbol{p} - \boldsymbol{y} = \mathbf{0}$ and $\boldsymbol{y} = \boldsymbol{p}$. We have justified the following result.

Lemma 12.2.1

Suppose $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Then there exists a unique vector $\boldsymbol{p} \in \text{col}(A^T)$ such that $A\boldsymbol{p} = \boldsymbol{b}$. Furthermore if $\boldsymbol{x} \in \mathbb{R}^n$ is any solution to $A\boldsymbol{x} = \boldsymbol{b}$ then

$$\boldsymbol{p} = \text{proj}_{\text{col}(A^T)}\boldsymbol{x}$$

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ with SVD

$$A = U_r \Sigma_r V_r^T = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

Let $\boldsymbol{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$. Then $A\boldsymbol{x} = \boldsymbol{b}$ has the solution $\boldsymbol{x} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$. To find the unique vector $\boldsymbol{p} \in \text{col}(A^T)$ such that $A\boldsymbol{p} = \boldsymbol{b}$ we calculate

$$\boldsymbol{p} = \text{proj}_{\text{col}(A^T)} \boldsymbol{x} = V_r V_r^T \boldsymbol{x} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 0 \\ 3/2 \end{bmatrix}$$

We calculate

$$\|\boldsymbol{p}\| = 3\sqrt{2}/2, \qquad \|\boldsymbol{x}\| = 3$$

and note that

$$\|\boldsymbol{p}\| \leq \|\boldsymbol{x}\| \tag{12.16}$$

As the next exercise shows, equation (12.16) is not a coincidence.

Exercise 12.9

Suppose $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Let $\boldsymbol{p}$ be the unique vector in $\text{col}(A^T)$ such that $A\boldsymbol{p} = \boldsymbol{b}$. Show that for any solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$ we have that

$$\|\boldsymbol{p}\| \leq \|\boldsymbol{x}\|$$

and furthermore $\|\boldsymbol{x}\| = \|\boldsymbol{p}\|$ if and only if $\boldsymbol{x} = \boldsymbol{p}$. Thus $\boldsymbol{p}$ is the *unique* solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length.

Suppose $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Suppose A has SVD given by $A = U_r \Sigma_r V_r^T$. Plugging the SVD into $A\boldsymbol{x} = \boldsymbol{b}$ where $\boldsymbol{x}$ is any solution gives

$$U_r \Sigma_r V_r^T \boldsymbol{x} = \boldsymbol{b} \tag{12.17}$$

Multiplying both sides on the left by $V_r \Sigma_r^{-1} U_r^T$ and using exercise 12.3 gives

$$V_r \Sigma_r^{-1} U_r^T \boldsymbol{b} = V_r \Sigma_r^{-1} U_r^T U_r \Sigma_r V_r^T \boldsymbol{x} = V_r V_r^T \boldsymbol{x} = \text{proj}_{\text{col}(A^T)} \boldsymbol{x} \tag{12.18}$$

Combining lemma 12.2.1, exercise 12.9, and equation 12.18 gives us the following result.

Lemma 12.2.2

Suppose $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Then $V_r \Sigma_r^{-1} U_r^T \boldsymbol{b}$ is the unique solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length.

Definition 12.1

Suppose A has SVD given by $A = U_r \Sigma_r V_r^T$. We define the pseudoinverse of A, denoted A^+, by

$$A^+ = V_r \Sigma_r^{-1} U_r^T$$

Consider again the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ with SVD given by

$$A = U_r \Sigma_r V_r^T = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

It is easy to check that $\boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ is in col(A) and thus $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. We calculate

$$A^+ = V_r \Sigma_r^{-1} U_r^T = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/4 & 0 & 1/4 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 1/4 & 0 & 1/4 \end{bmatrix}$$

By lemma 12.2.2 we have that

$$\boldsymbol{p} = A^+\boldsymbol{b} = \begin{bmatrix} 1/4 & 0 & 1/4 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 1/4 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/2 \\ 1/2 \\ 1 \end{bmatrix}$$

is the unique solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length.

Exercise 12.10

Let $A \in \mathbb{R}^{n \times n}$ and suppose that A is invertible. Prove that

$$A^+ = A^{-1}$$

Exercise 12.11

Find a reduced SVD for $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ and calculate A^+.

Suppose $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent. In this case $\boldsymbol{b}$ is not in col(A) so we transform to the consistent system

$$A\boldsymbol{x} = \text{proj}_{\text{col}(A)}\boldsymbol{b} \tag{12.19}$$

Recall that $\hat{\boldsymbol{x}}$ is a least square solution to $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\hat{\boldsymbol{x}}$ is a solution to equation (12.19). Thus, by lemma 12.2.2 we have that $A^+\text{proj}_{\text{col}(A)}\boldsymbol{b}$ is the unique solution to equation (12.19) of minimal length and hence the unique least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length. Using exercise 12.3 we have

$$A^+\text{proj}_{\text{col}(A)}\boldsymbol{b} = V_r\Sigma_r^{-1}U_r^TU_rU_r^T\boldsymbol{b} = V_r\Sigma_r^{-1}U_r^T\boldsymbol{b} = A^+\boldsymbol{b}$$

and we have justified the following result.

Theorem 12.2

Suppose $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$. Then $A^+\boldsymbol{b}$ is the unique least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length.

Consider again the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ with SVD given by

$$A = U_r\Sigma_rV_r^T = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

We recall the following calculations.

$$U_rU_r^T = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}, \qquad V_rV_r^T = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}, \qquad A^+ = \begin{bmatrix} 1/4 & 0 & 1/4 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 1/4 & 0 & 1/4 \end{bmatrix}$$

Let $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$. We first note that

$$\text{proj}_{\text{col}(A)}\boldsymbol{b} = U_r U_r^T \boldsymbol{b} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$$

and since $\boldsymbol{b} \neq \text{proj}_{\text{col}(A)}$, we have that $\boldsymbol{b}$ is not in $\text{col}(A)$. By theorem 12.2 we have that

$$\boldsymbol{x}^+ = A^+\boldsymbol{b} = \begin{bmatrix} 1/4 & 0 & 1/4 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 1/4 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1 \\ 1 \\ -1/2 \end{bmatrix}$$

is the unique least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length. Next we note that

$$\text{proj}_{\text{col}(A^T)}\boldsymbol{x}^+ = V_r V_r^T \boldsymbol{x}^+ = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} -1/2 \\ 1 \\ 1 \\ -1/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1 \\ 1 \\ -1/2 \end{bmatrix}$$

which verifies that $\boldsymbol{x}^+$ is in $\text{col}(A^T)$.

Exercise 12.12

Let $A \in \mathbb{R}^{3\times 3}$ and $\boldsymbol{b} \in \mathbb{R}^3$ where $A = U_r \Sigma_r V_r^T$ and

$$U_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}, \qquad \Sigma_r = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad V_r = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ -1/\sqrt{2} & 0 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

(1) Calculate A^+ and use it to calculate the unique least squares solution $\boldsymbol{x}^+$ to $A\boldsymbol{x} = \boldsymbol{b}$ of minimal length.
(2) Calculate $\text{proj}_{\text{col}(A^T)}\boldsymbol{x}^+$ and verify that $\boldsymbol{x}^+$ is in $\text{col}(A^T)$.

Exercise 12.13

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $\text{rank}(A) = r$. Prove that

$$\text{rank}(A^+) = r$$

Exercise 12.14

Let $A \in \mathbb{R}^{m\times n}$. Prove that for all $\boldsymbol{b} \in \mathbb{R}^m$ and all $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$AA^+\boldsymbol{b} = \text{proj}_{\text{col}(A)}\boldsymbol{b}, \qquad A^+A\boldsymbol{x} = \text{proj}_{\text{col}(A^T)}\boldsymbol{x}$$