Introduction to Linear Algebra

Math 107

Lecture Notes
Spring 2023

Ahmet Berkay Kebeci
akebeci@ku.edu.tr

Department of Mathematics
Koç University

Last update: May 12, 2025

Contents

CHAPTER 4

Vector Spaces

4.1. Vector Spaces and Subspaces

In section 1.3, we examined some algebraic properities of $\mathbb{R}^n$. (See page 37 on the handwritten notes.) We will define a vector space as a set with addition and scalar multiplication satisfying these algebraic properities.

Definition. *A* ***vector space*** *(over $\mathbb{R}$) consists of the following:*

(1) A non-empty set V of objects, called ***vectors****,*
(2) An operation called (vector) ***addition****, which associates with each pair of vectors $v, u \in V$, a vector $u + v \in V$, called the sum of u and v,*
(3) An operation called ***scalar multiplication****, which associates with each scalar $c \in \mathbb{R}$ and vector $v \in V$, a vector $c \cdot v \in V$, called the scalar multiple of v by c,*

such that for any $u, v, w \in V$ and for any scalars $c, d \in \mathbb{R}$

(i) $u + v = v + u$
(ii) $(u + v) + w = u + (v + w)$
(iii) There is a zero vector, denoted by 0, in V such that $u + 0 = 0$
(iv) For each $u \in V$, there is a vector $-u \in V$ such that $u + (-u) = 0$
(v) $c(u + v) = cu + cv$
(vi) $(c + d)u = cu + du$
(vii) $c(du) = (cd)u$
(viii) $1u = u$.

Here are some examples of vector spaces.

Example.

- $\mathbb{R}^n$, where $n \geq 1$ is an integer.
- The space $M_{m\times n}$ of $m \times n$ matrices, with matrix addition and scalar multiplication, where $m, n \geq 1$ are integers. (By theorem 2.1.)
- The space $\mathbb{P}_n$ of polynomials with real coefficients of degree at most n, where $n \geq 0$ is an integer. This consists of all polynomials of the form

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

where $a_i \in \mathbb{R}$. Addition and scalar multiplication is given by

$$(a_0 + a_1x + \ldots + a_nx^n) + (b_0 + b_1x + \ldots + b_nx^n) = (a_0 + b_0) + (a_1 + b_1)x + \ldots + (a_n + b_nx^n),$$
$$c(a_0 + a_1x + \ldots + a_nx^n) = ca_0 + ca_1x + \ldots + ca_nx^n.$$

- The space $\mathbb{P}$ (also denoted by $\mathbb{R}[x]$) of all polynomials with real coefficients.

- The space $\mathbb{R}^X$ of functions $X \longrightarrow \mathbb{R}$, where X is a set. Addition and scalar multiplication is given by
$$(f+g)(x) = f(x) + g(x),$$
$$(cf)(x) = cf(x),$$
where $c \in \mathbb{R}$ and $f, g : X \longrightarrow \mathbb{R}$ are functions.
- The space V^X of functions $X \longrightarrow V$, where X is a set and V is a vector space.
- The space $\mathbb{S}$ of all doubly infinite sequences of real numbers. This consists of the sequences of the form
$$\{y_k\}_{k\in\mathbb{Z}} = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots)$$
where $y_i \in \mathbb{R}$.
- $\mathbb{R}$, the set of real numbers, with usual addition and multiplication.
- $\mathbb{C}$, the set of complex numbers, with usual addition and multiplication.
$$(a+bi) + (c+di) = (a+c) + (b+d)i$$
$$c(a+bi) = ca + abi$$
where $a, b, c, d \in \mathbb{R}$.
- The space $\{0\}$ containing only the zero vector.

Observation.

- *If $u, v, w \in V$ such that $u + w = v + w$, then $u = v$. (Cancellation Law for Vector Addition)*
- *Zero vector in the axiom (iii) is unique.*
- *For any given $u \in V$, its additive inverse $-u$ in the axiom (iv) is unique.*

Proof.

- If $u + w = v + w$, then
$$\begin{aligned} u &= u + 0 = u + (w + (-w)) = (u + w) + (-w) \\ &= (v + w) + (-w) = v + (w + (-w)) = v + 0 = v. \end{aligned}$$
- If 0 and $0'$ are zero vectors, then $0 + 0' = 0$ since $0'$ is a zero vector and $0 + 0' = 0'$ since 0 is a zero vector. So $0' = 0 + 0' = 0$.
- If $u + v_1 = 0$ and $u + v_2 = 0$, then $u + v_1 = u + v_2$ and by cancellation law $v_1 = v_2$.

□

Observation. *For each $u \in V$ and $c \in \mathbb{R}$,*

- *$0u = 0$,*
- *$c0 = 0$,*
- *$-u = (-1)u$.*

Proof.

- Since $0u = (0+0)u = 0u + 0u$, we have $0 = 0u$.
- Since $c0 = c(0+0) = c0 + c0$, we have $0 = c0$.

- $u + (-1u) = 1u + (-1)u = (1 + (-1))u = 0u = 0$, so $-u = (-1u)$ by the uniqeness of additive inverse.

$\square$

Subspaces. Notice that $\mathbb{P}_3 \subseteq \mathbb{P}_4$ and both of them are vector spaces with the same addition and scalar multiplication. In this case, we will say that $\mathbb{P}_3$ is a subspace of $\mathbb{P}_4$.

A subset W of a vector space V will be called a *subspace* of V if W is also a vector space with the same operations of vector addition and scalar multiplication on V. It is easy to see that this definition is equivalent to the following definition.

Definition. *Let V be a vector space. A subset W of V is called a* ***subspace*** *if it satisfies the following three properties.*

(i) The zero vector of V is in W.
(ii) W is closed under vector addition. That is, for any $\mathbf{u}, \mathbf{v} \in W$, we have $\mathbf{u} + \mathbf{v} \in W$.
(iii) W is closed under multiplication by scalars. That is, for any $\mathbf{u} \in W$ and for any $c \in \mathbb{R}$, we have $c\mathbf{u} \in W$.

Example.

- For any vector space V, the set $\{0\}$ consisting of only the zero vector, is a subspace, called the *zero subspace*.
- Any vector space is a subspace of itself.
- A line in $\mathbb{R}^2$ through the origin is a subspace of $\mathbb{R}^2$.
- A line in $\mathbb{R}^3$ through the origin is a subspace of $\mathbb{R}^3$.
- A plane in $\mathbb{R}^3$ through the origin is a subspace of $\mathbb{R}^3$.
- In $\mathbb{R}^n$, the subset of vectors of the form $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ 0 \end{bmatrix}$ is a subspace.
- $\mathbb{P}_n$ is a subspace of $\mathbb{P}_m$, for any integers $m \geq n \geq 0$.
- $\mathbb{P}_n$ is a subspace of $\mathbb{P}$, for any integer $n \geq 0$.
- $\mathbb{P}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$.
- The diagonal $n \times n$ matrices form a subspace of $M_{m\times n}$.
- An $n \times n$ matrix A is called a *symmetric* matrix if $A^T = A$. The symmetric $n \times n$ matrices form a subspace of $M_{m\times n}$.
- The set
$$\{p(x) \in \mathbb{P} \mid p(5) = 0\}$$
is a subspace of $\mathbb{P}$.

Example.

- $\mathbb{R}^2$ is **not** a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$.
- A line in $\mathbb{R}^2$ not through the origin is not a subspace of $\mathbb{R}^2$ as it does not contain the zero vector of $\mathbb{R}^2$.

To check that if a subset is a subspace, we can check (ii) and (iii) in the definition above at once.

OBSERVATION. *A subset W of a vector space V is a subspace of V if and only if it satisfies the following properties.*

(i) The zero vector of V is in W.
(ii) For any $u, v \in W$, and for any $c \in \mathbb{R}$, we have $cu + v \in W$.

The proof of this is left as an exercise.

EXAMPLE. Let A be an $m \times n$ matrix. Then the solution set of the matrix equation $Ax = 0$ is a subspace of $\mathbb{R}^n$.

SOLUTION. 0 is in the solution set of $Ax = 0$. If u and v are two solutions of $Ax = 0$ and $c \in \mathbb{R}$, then $A(cu + v) = c(Au) + Av = c0 + 0 = 0$. So $cu + v$ is also a solution of $Ax = 0$.

A Subspace Spanned by a Set. Let V be a vector space and let $v_1, \dots, v_n \in V$.

DEFINITION. *We define* $\text{Span}\{v_1, \dots, v_n\}$ *as the set of linear combinations of* $v_1, \dots, v_n$.

$$\text{Span}\{v_1, \dots, v_n\} = \{c_1v_1 + \dots + c_nv_n \mid c_1, \dots, c_n \in \mathbb{R}\}.$$

THEOREM 4.1. $\text{Span}\{v_1, \dots, v_n\}$ *is a subspace of V.*

We call $\text{Span}\{v_1, \dots, v_n\}$ the **subspace spanned** (or **generated**) by $v_1, \dots, v_n$.

PROOF. We have $0 = 0v_1 + 0v_2 + \dots + 0v_n \in \text{Span}\{v_1, \dots, v_n\}$. Let $u, w \in \text{Span}\{v_1, \dots, v_n\}$ and let $r \in \mathbb{R}$. Then

$$\begin{aligned} u &= c_1v_1 + \dots + c_nv_n \\ w &= d_1v_1 + \dots + d_nv_n \end{aligned}$$

for some $c_i, d_i \in \mathbb{R}$. Then

$$u + rw = (c_1 + rd_1)v_1 + \dots (c_n + rd_n)v_n \in \text{Span}\{v_1, \dots, v_n\}.$$

$\square$

EXAMPLE. Show that

$$W = \left\{ \begin{bmatrix} a \\ -2a + b \\ 3a - 2b \\ b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^4$.

SOLUTION. An arbitrary vector in W has the form

$$\begin{bmatrix} a \\ -2a + b \\ 3a - 2b \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ -2 \\ 3 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix}$$

and this shows that

$$W = \left\{ a \begin{bmatrix} 1 \\ -2 \\ 3 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

$$= \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ -2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

Thus W is a subspace, by theorem 4.1.

4.2. Null Spaces, Column Spaces, Row Spaces, and Linear Transformations

Linear Transformations. In section 1.8, we defined linear transformations $\mathbb{R}^n \to \mathbb{R}^m$. Now, we will generalize this notion to maps between vector spaces.

DEFINITION. *A* ***linear transformation*** *(or* ***linear map****) from a vector space V into a vector space W is a function $T : V \to W$ such that*

$$T(cu + v) = cT(u) + T(v)$$

for all $u, v \in V$ and for all $c \in \mathbb{R}$.

EXAMPLE. The derivative $D : \mathbb{P} \to \mathbb{P}$,

$$D(a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_nx^n) = a_1 + 2a_2x + 3a_3x^2 + \ldots + na_nx^{n-1}$$

is a linear transformation.

EXAMPLE. The function $T : \mathbb{P}_n \to \mathbb{R}^{n+1}$ given by

$$T(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

is a linear transformation.

- Recall that **range** of a linear transformation $T : V \to W$ is the subset
$$\operatorname{Range}(T) = \{T(u) \in W \mid u \in V\}$$
of W.
- We define the **kernel** (or **null space**) of a linear transformation $T : V \to W$ as the subset
$$\operatorname{Ker}(T) = \{u \in V \mid T(u) = 0\}$$
of V.

OBSERVATION. *Let $T : V \to W$ be a linear transformation.*
(i) $T(0) = 0$.
(ii) Range of T is a subspace of W.

(iii) Kernel of T is a subspace of V.
(iv) T is onto if and only if range of T is W.
(v) T is one-to-one if and only if kernel of T is $\{0\}$.

Proof.

(i) $T(0) = T(0+0) = T(0) + T(0)$, so $0 = T(0)$.

(ii) 0 is in the range, since $T(0) = 0$. Let w_1, w_2 be in the range of T and let $c \in \mathbb{R}$. Then $w_1 = T(v_1)$ and $w_2 = T(v_2)$ for some $v_1, v_2 \in V$. Then

$$T(cv_1 + v_2) = cT(v_1) + T(v_2) = cw_1 + w_2,$$

so $cw_1 + w_2$ is in the range of T. Thus range of T is a subspace of W.

(iii) 0 is in the kernel, since $T(0) = 0$. Let v_1, v_2 be in the kernel of T and let $c \in \mathbb{R}$. Then $T(v_1) = 0$ and $T(v_2) = 0$. Then

$$T(cv_1 + v_2) = cT(v_1) + T(v_2) = c0 + 0 = 0$$

so $cv_1 + v_2$ is in the kernel of T. Thus kernel of T is a subspace of V.

(iv) By definition.

(v) If T is one-to-one and if $T(u) = 0$, then $T(u) = 0 = T(0)$ and we have $u = 0$. If kernel of T is $\{0\}$ and if $T(u) = T(v)$, then $T(u - v) = T(u) - T(v) = 0$ and $T(u) = T(v)$.

$\square$

Example. Let

$$T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$$
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + x_2 - 5x_3 \\ x_1 - 3x_2 - x_3 \\ x_1 - x_2 + 6x_3 \end{bmatrix}$$

Then T is linear, with the standard matrix

$$A = \begin{bmatrix} 1 & 2 & -5 \\ 1 & -3 & -1 \\ 1 & -2 & 6 \end{bmatrix}.$$

Kernel of T is the set of all vectors $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ satisfying

$$\begin{bmatrix} x_1 + x_2 - 5x_3 \\ x_1 - 3x_2 - x_3 \\ x_1 - x_2 + 6x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

that is the solution set of linear system

$$\begin{aligned} x_1 + x_2 - 5x_3 &= 0 \\ x_1 - 3x_2 - x_3 &= 0 \\ x_1 - x_2 + 6x_3 &= 0. \end{aligned}$$

This is also the solution set of the homogeneous equation $Ax = 0$.
Range of T is the set of all vectors of the form

$$\begin{bmatrix} x_1 + x_2 - 5x_3 \\ x_1 - 3x_2 - x_3 \\ x_1 - x_2 + 6x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} -5 \\ -1 \\ 6 \end{bmatrix},$$

that is the span of

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} -5 \\ -1 \\ 6 \end{bmatrix} \right\}.$$

So this is given by the span of the columns of A.

The Null Space and the Column Space of a Matrix.

DEFINITION. *Let A be an $m \times n$ matrix.*

- *The **null space** of A, written as $\operatorname{Nul} A$, is defined as kernel of the linear transformation $x \mapsto Ax$, that is, the set of all solutions of the homogeneous equation $Ax = 0$.*
$$\operatorname{Nul} A = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$
- *The **column space** of A, written as $\operatorname{Col} A$, is defined as range of the linear transformation $x \mapsto Ax$, that is, the set of all linear combinations of the columns of A. If $A = \begin{bmatrix} a_1 & \dots & A_n \end{bmatrix}$, then*
$$\begin{aligned} \operatorname{Col} A &= \{b \in \mathbb{R}^m \mid b = Ax \text{ for some } x \in \mathbb{R}^n\} \\ &= \operatorname{Span}\{a_1, \dots, a_n\}. \end{aligned}$$

EXAMPLE. Let $A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}$. Is $u = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$ in the null space of A.

SOLUTION. We have $Au = 0$, so $u \in \operatorname{Nul} A$.

We have already proved the following results in the observation above.

THEOREM 4.2. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $Ax = 0$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.*

THEOREM 4.3. *The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.*

EXAMPLE. Let W be the set of all vectors $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathbb{R}^4$ satisfying $a - 2b + c = 5d$ and $c - a = b$. Show that W is a subspace of $\mathbb{R}^4$.

SOLUTION. W is the solution set of the homogeneous linear system

$$\begin{aligned} a - 2b + c - 5d &= 0 \\ -a - b + c &= 0. \end{aligned}$$

Thus, W is a subspace of $\mathbb{R}^4$.

- Solving the equation $Ax = 0$ gives an explicit description of $\operatorname{Nul} A$.

EXAMPLE. Find a spanning set for the null space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION. We need to solve $Ax = 0$.

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

This corresponds to

$$\begin{aligned} x_1 - 2x_2 - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0. \end{aligned}$$

Its solutions are

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}.$$

Thus

$$\operatorname{Nul} A = \operatorname{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

EXAMPLE. Find a spanning set for the column space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION.

$$\operatorname{Col} A = \operatorname{Span} \left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \right\}.$$

EXAMPLE. Find a matrix A such that $\operatorname{Col} A = W$, where

$$W = \left\{ \begin{bmatrix} 7a - 6b \\ a + b \\ 3a \end{bmatrix} \mid a, b \in \mathbb{R} \right\}.$$

SOLUTION. Since

$$W = \left\{ a \begin{bmatrix} 7 \\ 1 \\ 3 \end{bmatrix} + b \begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix} \mid a, b \in \mathbb{R} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} 7 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix} \right\}$$

we have $\operatorname{Col} A = W$, for $A = \begin{bmatrix} 7 & -6 \\ 1 & 1 \\ 3 & 0 \end{bmatrix}$.

EXAMPLE. Let $A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 9 & 17 \\ 3 & 12 & 23 \\ 4 & 17 & 27 \end{bmatrix}$. Is $b = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 6 \end{bmatrix}$ in the column space of A?

SOLUTION. $b \in \operatorname{Col} A$ if and only if $Ax = b$ is consistent.

$$\begin{bmatrix} A & b \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 5 & 1 \\ 0 & 1 & 7 & -2 \\ 0 & 0 & 8 & -1 \\ 0 & 0 & 0 & -4 \end{bmatrix}.$$

This represents an inconsistent system. Thus $b \notin \operatorname{Col} A$.

OBSERVATION. *Let A be an $m \times n$ matrix.*

- *The following are equivalent.*
  - *(i)* $\operatorname{Nul} A = \{0\}$.
  - *(ii) The equation $Ax = 0$ has only the trivial solution.*
  - *(iii) The linear transformation $x \mapsto Ax$ is one-to-one.*
- *The following are equivalent.*
  - *(i)* $\operatorname{Col} A = \mathbb{R}^m$.
  - *(ii) The equation $Ax = b$ has a solution for every $b \in \mathbb{R}^m$.*
  - *(iii) The linear transformation $x \mapsto Ax$ is onto.*

The Row Space. If A is an $m \times n$ matrix, then A is of the form

$$A = \begin{bmatrix} \operatorname{row}_1(A) \\ \operatorname{row}_2(A) \\ \vdots \\ \operatorname{row}_m(A) \end{bmatrix}.$$

We can see each row as a vector in $\mathbb{R}^n$. ($(\operatorname{row}_i(A))^T \in \mathbb{R}^n$.)

The *row space* of A, written as $\operatorname{Row} A$, is defined as the set of all linear combinations of the rows of A.

$$\operatorname{Row} A = \operatorname{Span}\{(\operatorname{row}_1(A))^T, \dots, (\operatorname{row}_m(A))^T\}.$$

In other words $\operatorname{Row} A = \operatorname{Col} A^T$. Then $\operatorname{Row} A$ is a subspace of $\mathbb{R}^n$.

4.3. Linearly Independent Sets; Bases

Linear Independence. Recall the definition of linear Independence and dependence in $\mathbb{R}^n$. The same definition also generalizes to an arbitrary vector space.

DEFINITION. *Let V be a vector space. A subset $\{v_1, \ldots, v_n\}$ of V is called **linearly dependent** if there exist scalars $c_1, \ldots, c_n \in \mathbb{R}$, not all of which are 0, such that*

$$c_1v_1 + \ldots + c_nv_n = 0.$$

*It is called **linearly independent** otherwise, that is if the vector equation*

$$x_1v_1 + \ldots + x_nv_n = 0$$

has only the trivial solution $x_1 = 0, \ldots, x_n = 0$.

Just as in $\mathbb{R}^n$,

- $\{v\}$ is linearly independent $\iff$ $v \neq 0$
- $\{v_1, v_2\}$ is linearly independent $\iff$ neither of v_1 and v_2 is a multiple of the other

The following theorem has the same proof as theorem 1.7. and the observation after that.

THEOREM 4.4. *Let $S = \{v_1, \ldots, v_n\}$ be a subset of a vector space V such that $n > 1$ and $v_1 \neq 0$. Then, S is linearly dependent if and only if we have*

$$v_j \in \mathsf{Span}\{v_1, \ldots, v_{j-1}\}$$

for some $j > 1$.

EXAMPLE. Let $p_1(x) = 1$, $p_2(x) = x, p_3(x) = x^2$ and $p_4(x) = 2 - 3x + x^2$. Then $\{p_1, p_2, p_3, p_4\}$ is linearly dependent in $\mathbb{P}$ because

$$p_4 = 2p_1 - 3p_2 + p_3.$$

OBSERVATION. *Let $T : V \longrightarrow W$ be a linear transformation.*

(i) $\{v_1, \ldots, v_n\}$ is linearly dependent in V $\implies$ $\{T(v_1), \ldots, T(v_n)\}$ is linearly dependent in W.
(ii) $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent in W $\implies$ $\{v_1, \ldots, v_n\}$ is linearly independent in V.
(iii) T is one-to-one and $\{v_1, \ldots, v_n\}$ is linearly independent in V $\implies$ $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent in W.

PROOF.

(i) If $\{v_1, \ldots, v_n\}$ is linearly dependent, then there exist $c_1, \ldots, c_n \in \mathbb{R}$, not all of which are 0, such that

$$c_1v_1 + \ldots + c_nv_n = 0.$$

Then,

$$\begin{aligned} c_1T(v_1) + \ldots c_nT(v_n) &= T(c_1v_1 + \ldots + c_nv_n) \\ &= T(0) \\ &= 0. \end{aligned}$$

(ii) Follows by (i).
(iii) Consider the equation
$$x_1 T(v_1) + \ldots + x_n T(v_n) = 0.$$
This is the same as
$$T(x_1 v_1 + \ldots x_n v_n) = 0.$$
If T is one-to-one, then
$$x_1 v_1 + \ldots + x_n v_n = 0.$$
If $\{v_1, \ldots, v_n\}$ is linearly independent, then $x_1 = 0, \ldots, x_n = 0$.

$\square$

Remark. We may have $\{v_1, \ldots, v_n\}$ is linearly independent but $\{T(v_1), \ldots, T(v_n)\}$ is linearly dependent, when T is not one-to-one. The most basic example of that is when $T = 0$ (mapping everything to 0). Also consider the following example.

Example. Let $T : \mathbb{R}^5 \longrightarrow \mathbb{R}^3$ be the linear transformation whose standard matrix is
$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$
We have $T(e_1) = \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}$ and $T(e_2) = \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix} = -2T(e_1)$. So $\{e_1, e_2\}$ is linearly independent in $\mathbb{R}^5$ but $\{T(e_1), T(e_2)\}$ is linearly dependent in $\mathbb{R}^3$.

Observation.

- *Any set which contains a linearly dependent set is linearly dependent.*
- *Any subset of a linearly independent set is linearly independent.*
- *Any set which contains the* 0 *vector is linearly dependent.*

The proof of this is left as an exercise.

Basis. Consider the vectors
$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$
from $\mathbb{R}^3$. Any $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbb{R}^3$ can be written as a linear combination of e_1, e_2, e_3 in the unique way
$$b = b_1 e_1 + b_2 e_2 + b_3 e_3.$$
Also let
$$v_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Any $b \in \mathbb{R}^3$ can be written as a linear combination of v_1, v_2, v_3 because $\operatorname{Span}\{v_1, v_2, v_3\} = \mathbb{R}^3$. This representation is also unique as the vector equation

$$x_1 v_1 + x_2 v_2 + x_3 v_3 = b$$

has the unique solution if it is consistent. This is because $\{v_1, v_2, v_3\}$ is linearly independent.

Definition. *Let V be a vector space. A set of vectors $\mathcal{B}$ in V is called a* **basis** *for V if*

(i) $\mathcal{B}$ is linearly independent, and
(ii) $\operatorname{Span}\mathcal{B} = V$.

Example.

- $\{e_1, e_2, \dots, e_n\} \subseteq \mathbb{R}^n$ is a basis for $\mathbb{R}^n$. This is called the *standard basis* for $\mathbb{R}^n$.
- $\{1, x, x^2, \dots, x^n\}$ is a basis for $\mathbb{P}_n$. This is called the *standard basis* for $\mathbb{P}^n$.

• Let A be an invertible $n \times n$ matrix. Then the columns of A form a basis for $\mathbb{R}^n$ because they are linearly independent and they span $\mathbb{R}^n$, by the Invertible Matrix Theorem (theorem 2.8).

• Let $\{v_1, v_2, \dots, v_m\}$ be a basis for $\mathbb{R}^n$. Since it is linearly independent, $m \le n$, by theorem 1.8. Since it spans $\mathbb{R}^n$, we have $n \le m$. Thus $m = n$. Then the matrix $\begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$ is invertible by the Invertible Matrix Theorem (theorem 2.8), since its columns are linearly independent.

To sum up:

Observation. *$\{v_1, \dots, v_m\}$ is a basis for $\mathbb{R}^n$ $\iff$ $n = m$ and $\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \sim I_n$.*

Example. $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is a basis for $\mathbb{R}^3$ since $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}$ is invertible.

Spanning set theorem.

Example. Let

$$v_1 = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}, v_2 = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$$

and

$$H = \operatorname{Span}\{v_1, v_2, v_3\}.$$

Notice that $V_3 = v_1 + 2v_2$. Then $\operatorname{Span}\{v_1, v_2\} = \operatorname{Span}\{v_1, v_2, v_3\} = H$. Clearly, $\{v_1, v_2\}$ is linearly independent. Thus $\{v_1, v_2\}$ is a basis for H.

Theorem 4.5 (The Spanning Set Theorem). *Let V be a vector space, $S = \{v_1, \dots, v_m\} \subseteq V$ and $H = \operatorname{Span} S$.*

(a) If $v_k \in \operatorname{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\}$, then $\operatorname{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\} = H$.
(b) If $H \neq \{0\}$, then a subset of S is a basis for H.

PROOF. (a) $\underline{\text{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\} \subseteq H}$: If $u \in \text{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\}$, then for some $c_i \in \mathbb{R}$,

$$\begin{aligned} u &= c_1 v_1 + \dots + c_{k-1} v_{k-1} + c_{k+1} v_{k+1} + \dots + c_m v_m \\ &= c_1 v_1 + \dots + c_{k-1} v_{k-1} + 0 v_k + c_{k+1} v_{k+1} + \dots + c_m v_m \\ &\in \text{Span}\{v_1, \dots, v_m\} = H. \end{aligned}$$

$\underline{H \subseteq \text{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\}}$: Since $v_k \in \text{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\}$,

$$v_k = d_1 v_1 + \dots + d_{k-1} v_{k-1} + d_{k+1} v_{k+1} + \dots + d_m v_m$$

for some $d_i \in \mathbb{R}$. If $w \in H = \text{Span}\{v_1, \dots, v_m\}$, then for some $b_i \in \mathbb{R}$,

$$\begin{aligned} w &= b_1 v_1 + \dots + b_{k-1} b_{k-1} + + b_k v_k + b_{k+1} v_{k+1} + \dots + b_m v_m \\ &= (b_1 + b_k d_1) v_1 + \dots (b_{k-1} + b_k d_{k-1}) v_{k-1} + (b_{k+1} + b_k d_{k+1}) v_{k+1} + \dots + (b_m + b_k d_m) v_m \\ &\in \text{Span}\{v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m\}. \end{aligned}$$

(b) If S is linearly independent, then it is already a basis for H. Otherwise, one of the vectors in S depends on the others and can be deleted, by part (a). So long as there are two or more vectors in the spanning set, we can repeat this process until the spanning set is linearly independent and hence is a basis for H. If the spanning set is eventually reduced to one vector, that vector will be nonzero (and hence linearly independent) because $H \neq \{0\}$.

□

• A basis is a spanning set that is as small as possible. A basis is also a linearly independent set that is as large as possible.

EXAMPLE.

- $\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix} \right\}$ is linearly independent but does not span $\mathbb{R}^3$.
- $\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 11 \\ 13 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$.
- $\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 11 \\ 13 \end{bmatrix}, \begin{bmatrix} 17 \\ 19 \\ 23 \end{bmatrix} \right\}$ spans $\mathbb{R}^3$ but is linearly dependent.

Bases for Nul A, Col A, and Row A.

EXAMPLE. Find a basis for Nul B, Col B and Row B, where

$$B = \begin{bmatrix} b_1 & b_2 & \dots & b_7 \end{bmatrix} = \begin{bmatrix} 1 & 7 & 0 & 0 & -2 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 & 0 & -6 \\ 0 & 0 & 0 & 1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

SOLUTION. Note that B is already in the reduced echelon form.

- <u>Nul B</u>: Solving $Bx = 0$, we have

$$\begin{aligned} x_1 &= -7x_2 + 2x_5 + x_7 \\ x_2 &\text{ is free} \\ x_3 &= -3x_5 + 6x_7 \\ x_4 &= -4x_5 \\ x_5 &\text{ is free} \\ x_6 &= 0 \\ x_7 &\text{ is free.} \end{aligned}$$

So

$$\operatorname{Nul} B = \operatorname{Span}\left\{ \begin{bmatrix} -7 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \operatorname{Span}\{v_1, v_2, v_3\}.$$

By looking at 2nd, 5th and 7th entries of v_1, v_2, v_3, we can see that they are linearly independent. Hence

$$\{v_1, v_2, v_3\} = \left\{ \begin{bmatrix} -7 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\operatorname{Nul} B$.

- <u>Col B</u>: By definition, $\operatorname{Col} B = \operatorname{Span}\{b_1, \dots b_7\}$. Notice that

$$\begin{aligned} b_2 &= 7b_1 \\ b_5 &= -2b_1 + 3b_3 + 4b_4 \\ b_7 &= -b_1 - 6b_3. \end{aligned}$$

So $\operatorname{Col} B = \operatorname{Span}\{b_1, \dots b_7\} = \operatorname{Span}\{b_1, b_3, b_4, b_6\}$. It is easy to see that b_1, b_3, b_4, b_6 are linearly independent. Thus

$$\{b_1, b_3, b_4, b_6\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $\operatorname{Col} B$.

- Row B: It is easy to see that

  $$\{(1,7,0,0,-2,0,-1),(0,0,1,0,3,0,-6),(0,0,0,1,4,0,0),(0,0,0,0,0,1,0)\}$$

  is a basis for Row B.

EXAMPLE. Find a basis for Nul A, Col A and Row A, where

$$A = \begin{bmatrix} a_1 & a_2 & \dots & a_7 \end{bmatrix} = \begin{bmatrix} 1 & 7 & 0 & 0 & -2 & 1 & -1 \\ 3 & 21 & 1 & 0 & -3 & -1 & -9 \\ -6 & -42 & 1 & 1 & 19 & -4 & 0 \\ -2 & -14 & 0 & -1 & 0 & 3 & 2 \\ 0 & 0 & -4 & 3 & 0 & 5 & 24 \end{bmatrix}.$$

SOLUTION. Via row reduction, it can be shown that A is is row equivalent to the matrix B in the example above.

- Nul A: Since $A \sim B$, the solution set of $Ax = 0$ is the same as the solution set of $Bx = 0$. So Nul A = Nul B. Hence

  $$\left\{ \begin{bmatrix} -7 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

  is a basis for Nul A.
- Col A: Row reduction protects all the linear relations between the column vectors. So we also have,

  $$\begin{aligned} a_2 &= 7a_1 \\ a_5 &= -2a_1 + 3a_3 + 4a_4 \\ a_7 &= -a_1 - 6a_3 \end{aligned}$$

  since $A \sim B$. Therefore, Col A = $\mathrm{Span}\{a_1, \dots a_7\} = \mathrm{Span}\{a_1, a_3, a_4, a_6\}$. Similarly, since b_1, b_3, b_4, b_6 are linearly independent, a_1, a_3, a_4, a_6 are linearly independent. Thus

  $$\{a_1, a_3, a_4, a_6\} = \left\{ \begin{bmatrix} 1 \\ 3 \\ -6 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -4 \\ 3 \\ 5 \end{bmatrix} \right\}$$

  is a basis for Col A.
- Row A: B can be obtained from A by row operations since $A \sim B$ and therefore the rows of B are linear combinations of the rows of A. So, Row $B \subseteq$ Row A. Similarly, A can be obtained from B by row operations, so Row $A \subseteq$ Row B. Therefore Row A = Row B. Thus

  $$\{(1,7,0,0,-2,0,-1),(0,0,1,0,3,0,-6),(0,0,0,1,4,0,0),(0,0,0,0,0,1,0)\}$$

is a basis for Row A.

Generalizing the discussion in the examples above, we have the following results.

OBSERVATION. *If $A \sim B$, then* $\operatorname{Nul} A = \operatorname{Nul} B$.

THEOREM 4.6. *The pivot columns of a matrix A form a basis for* $\operatorname{Col} A$.

THEOREM 4.7. *If $A \sim B$, then* $\operatorname{Row} A = \operatorname{Row} B$. *If B is in echelon form, the nonzero rows of B form a basis for* $\operatorname{Row} A = \operatorname{Row} B$.

4.4. Coordinate Systems

THEOREM 4.8 (The Unique Representation Theorem). *Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis for a vector space V. Then for each $u \in V$, there exists a unique set of scalars $c_1, \ldots, c_n \in \mathbb{R}$ such that*

$$u = c_1 b_1 + \ldots + c_n b_n.$$

PROOF. Since $\operatorname{Span} \mathcal{B} = V$, there exists such scalars. If also

$$u = d_1 b_1 + \ldots + d_n b_n$$

for some $d_1, \ldots, d_n \in V$, then

$$0 = (c_1 - d_1) b_1 + \ldots + (c_n - d_n) b_n.$$

Since $\mathcal{B}$ is linearly independent, $c_1 - d_1 = 0$, $c_2 - d_2 = 0$, …, $c_n - d_n = 0$. Thus $c_1 = d_1$, $c_2 = d_2$, …, $c_n = d_n$. ☐

DEFINITION.

- *Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis for a vector space V and $u \in V$. The* ***coordinates of u relative to the basis $\mathcal{B}$*** *(or the $\mathcal{B}$-**coordinates** of u) are the unique numbers $c_1, \ldots, c_n \in \mathbb{R}$ such that*

$$u = c_1 b_1 + \ldots + c_n b_n.$$

- *In this case we call*

$$[u]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

*$\mathcal{B}$-**coordinate vector** of u.*

EXAMPLE. Let $b_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $b_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then $\mathcal{B} = \{b_1, b_2\}$ is a basis for $\mathbb{R}^2$. If an $u \in \mathbb{R}^2$ has the coordinate vector $[u]_{\mathcal{B}} = \begin{bmatrix} 7 \\ -3 \end{bmatrix}$, then

$$u = 7b_1 + (-3)b_2 = 7 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}.$$

EXAMPLE. The entries in the vector $u = \begin{bmatrix} 4 \\ 10 \end{bmatrix}$ are the coordinates of u relative to the standard basis $\varepsilon = \{e_1, e_2\}$, since

$$\begin{bmatrix} 4 \\ 10 \end{bmatrix} = 4e_1 + 10e_2.$$

So, if $\varepsilon = \{e_1, e_2\}$, then $[u]_\varepsilon = u$.

EXAMPLE. Let $v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $v_2 = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$ and $H = \operatorname{Span}\{v_1, v_2\}$. Let $u = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}$. Then $\mathcal{B} = \{v_1, v_2\}$ is a basis for H. Is $u \in H$? If it is, find $[u]_\mathcal{B}$.

SOLUTION. We need to solve $x_1 v_1 + x_2 v_2 = u$. This corresponds to the following augmented matrix.

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

So, $u \in H$ and $u = v_1 + 2v_2$. Thus $[u]_\mathcal{B} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

EXAMPLE. Let $\mathcal{B} = \{1, x, x^2, x^3\} \subseteq \mathbb{P}_3$. Then $\mathcal{B}$ is a basis for $\mathbb{P}_3$. Consider

$$p = 3 - 2x + x^3.$$

Then,

$$p = 3 \cdot 1 + (-2) \cdot x + 0 \cdot x^2 + 1 \cdot x^3$$

and $[p]_\mathcal{B} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \end{bmatrix}$.

Coordinate mapping.

DEFINITION. *Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis for a vector space V. The mapping*

$$\begin{aligned} V &\longrightarrow \mathbb{R}^n \\ u &\longmapsto [u]_\mathcal{B} \end{aligned}$$

is called the ***coordinate mapping*** *(**determined by** $\mathcal{B}$).*

DEFINITION. *A one-to-one, onto linear transformation is called an* ***isomorphism***.

THEOREM 4.9. *The coordinate mapping is an isomorphism.*

PROOF. Linearity: Let $u, v \in V$ and $r \in \mathbb{R}$. Then

$$\begin{aligned} u &= c_1 b_1 + \ldots + c_n b_n \\ v &= d_1 b_1 + \ldots + d_n b_n \end{aligned}$$

for some unique $c_i, d_i \in \mathbb{R}$ and

$$[u]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}, \quad [v]_{\mathcal{B}} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}.$$

Then,

$$ru + v = (rc_1 + d_1)b_1 + \ldots + (rc_n + d_n)b_n.$$

So

$$[ru + v]_{\mathcal{B}} = \begin{bmatrix} rc_1 + d_1 \\ \vdots \\ rc_n + d_n \end{bmatrix} = r \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = r[u]_{\mathcal{B}} + [v]_{\mathcal{B}}.$$

Thus, the coordinate mapping is linear.

<u>One-to-one</u>: Suppose $[u]_{\mathcal{B}} = [v]_{\mathcal{B}}$ for some $u, v \in V$. Call

$$[u]_{\mathcal{B}} = [v]_{\mathcal{B}} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

Then,

$$u = h_1 b_1 + \ldots + h_n b_n = v.$$

So, the coordinate mapping is one-to-one.

<u>Onto</u>: Let $\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \in \mathbb{R}^n$ be an arbitrary vector. Consider

$$u = h_1 b_1 + \ldots + h_n b_n \in V.$$

Then $[u]_{\mathcal{B}} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}$. Hence, the coordinate mapping is onto. □

Example.

- $\mathcal{B} = \{1, x, x^2, x^3\}$ is the standard basis for $\mathbb{P}_3$. If $p = a_0 + a_1 x + a_2 x^2 + a_3 x^3$, then

$$[p]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

The map

$$\begin{aligned} \mathbb{P}_2 &\longrightarrow \mathbb{R}^3 \\ p &\longmapsto [p]_{\mathcal{B}} \end{aligned}$$

is an isomorphism.

- In general, $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$ is the standard basis for $\mathbb{P}_n$. An arbitrary element of $\mathbb{P}_n$ is of the form

$$p = a_o + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

and

$$[p]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Similarly,

$$\begin{aligned} \mathbb{P}_n &\longrightarrow \mathbb{R}^{n+1} \\ a_o + a_1 x + a_2 x^2 + \ldots + a_n x^n &\longmapsto \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \end{aligned}$$

is an isomorphism.

Observation. *If $T : V \longrightarrow W$ is an isomorphism,*

$$\{v_1, \ldots, v_n\} \text{ is linearly independent in } V \iff \{T(v_1), \ldots, T(v_n)\} \text{ is linearly independent in } W.$$

Example. Is $\{1 + 2x^2, 4 + x + 5x^2, 3 + 2x\}$ linearly independent in $\mathbb{P}_2$?

Solution. $\{1 + 2x^2, 4 + x + 5x^2, 3 + 2x\}$ is linearly independent in $\mathbb{P}_2$ if and only if $\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \right\}$ is linearly independent in $\mathbb{R}^3$. To check that we need to solve the homogeneous system represented by the following augmented matrix.

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

So, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \right\}$ is linearly dependent. Thus $\{1 + 2x^2, 4 + x + 5x^2, 3 + 2x\}$ is linearly dependent.

Coordinates in $\mathbb{R}^n$. Let $\mathcal{B}$ be a basis for $\mathbb{R}^n$. The coordinate mapping

$$\begin{aligned} \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ u &\longmapsto [u]_{\mathcal{B}} \end{aligned}$$

corresponds to coordinate change.

EXAMPLE. Consider $\mathbb{R}^2$ with a basis $\mathcal{B} = \{b_1, b_2\}$, where $b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, b_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Let $v = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$. Solving

$$\begin{bmatrix} 2 & -1 & 4 \\ 1 & 1 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

gives $v = 3b_1 + 2b_2$. So

$$[v]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Here we have

$$\begin{aligned} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} &= \begin{bmatrix} 4 \\ 5 \end{bmatrix} \\ \implies \begin{bmatrix} b_1 & b_2 \end{bmatrix} [v]_{\mathcal{B}} &= v. \end{aligned}$$

In general, let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis for $\mathbb{R}^n$. Let $v \in \mathbb{R}^n$. Call $[v]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$.

$$\begin{aligned} &\implies v = c_1 b_1 + \dots c_n b_n \\ &\implies v = \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\ &\implies v = P_{\mathcal{B}}[v]_{\mathcal{B}}. \end{aligned}$$

We call $\begin{bmatrix} b_1 & \dots & b_n \end{bmatrix}$ the *change-of-coordinates matrix* from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$ and denote

$$P_{\mathcal{B}} = \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix}.$$

OBSERVATION. $v = P_{\mathcal{B}}[v]_{\mathcal{B}}$.

Since $\mathcal{B}$ is a basis for $\mathbb{R}^n$, the matrix $P_{\mathcal{B}}$ is invertible.

$$\implies P_{\mathcal{B}}^{-1} v = [v]_{\mathcal{B}}.$$

OBSERVATION. *$P_{\mathcal{B}}^{-1}$ is the standard matrix for the coordinate mapping*

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R}^n \\ u &\mapsto [u]_{\mathcal{B}}. \end{aligned}$$

4.5. The Dimension of a Vector Space

THEOREM 4.10. *If a vector space V has a basis*

$$\mathcal{B} = \{b_1, \dots, b_n\}$$

then any subset S of V with $|S| > n$ is linearity dependent.

Proof. Suppose $S = \{v_1, \ldots, v_m\}$ with $m > n$. Then

$$\implies \{[v_1]_{\mathcal{B}}, \ldots, [v_m]_{\mathcal{B}}\}$$

is linearly independent in $\mathbb{R}^n$ because $m > n$.

$$\implies \{v_1, \ldots, v_m\}$$

is linearity dependent in V. □

Definition. *If a vector space V is spanned by a finite set, then V is called* ***finite-dimensional****. Otherwise V is called* ***infinite-dimensional****.*

Example.

- $\mathbb{R}^n$ is finite-dimensional.
- $\mathbb{P}$ is infinite-dimensional.

Theorem 4.11. *If V is a finite-dimensional vector space, then any two bases of V has the same (finite) number of elements.*

Proof. Since V is spanned by a finite set, V has a finite basis.

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases for V. Since $\mathcal{B}_2$ is linearly independent and $\mathcal{B}_1$ is a basis, by theorem 4.10, $|\mathcal{B}_2| \leq |\mathcal{B}_1|$. Similarly, by theorem 4.10, we have $|\mathcal{B}_1| \leq |\mathcal{B}_2|$. So $|\mathcal{B}_1| = |\mathcal{B}_2|$. □

Definition. *Let V be a finite-dimensional vector space. The* ***dimension*** *of V, written as $\dim V$, is the number of vectors in a basis for V.*

Note that $\dim\{0\}$ is defined to be 0.

Example.

- $\dim \mathbb{R}^n = n$
- $\dim \mathbb{P}_n = n + 1$

Example. Let

$$H = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix} \right\}.$$

Since $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix}$ are not multiple of each other, they are linearly independent. Therefore $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix} \right\}$ is a basis for H. Thus $\dim H = 2$.

EXAMPLE. Let $v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $v_2 = \begin{bmatrix} -1 \\ 0 \\ 5 \\ -4 \end{bmatrix}$, $v_3 = \begin{bmatrix} -1 \\ 2 \\ 13 \\ -2 \end{bmatrix}$, $v_4 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}$ and $H = \operatorname{Span}\{v_1, v_2, v_3, v_4\}$. Then $H = \operatorname{Col} A$, where $A = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix}$.

$$A = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 2 & 0 & 2 & 2 \\ 3 & 5 & 13 & 3 \\ 4 & -4 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & -1 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus $\{v_1, v_2, v_3\}$ is a basis for H and therefore $\dim H = 3$.

Subspaces of a Finite-Dimensional Space.

THEOREM 4.12. *Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and*

$$\dim H \leq \dim V.$$

PROOF. If $H = \{0\}$, then $\dim H = 0 \leq \dim V$.

Suppose $H \neq \{0\}$ and let $S = \{v_1, \dots, v_m\}$ be a linearly independent subset of H. If $\operatorname{Span} S = H$, then S is a basis for H. Otherwise, there is some $v_{m+1} \in H$ such that $v_{m+1} \notin \operatorname{Span} S$. Then $\{v_1, \dots, v_m, v_{m+1}\}$ is linearly independent. Again if this spans H, then it is a basis. If not, then we can again expend it.

So long as the new set does not span H, we can continue this process of expanding S to a larger linearly independent set in H. But the number of vectors in a linearly independent expansion of S can never exceed the dimension of V, by theorem 4.10. So eventually the expansion of S will span H and hence will be a basis for H, and $\dim H \leq \dim V$. □

EXAMPLE. The subspaces of $\mathbb{R}^3$:

- 0-*dimensional subspaces*: Only $\{0\}$.
- 1-*dimensional subspaces*: $\operatorname{Span}\{v\}$, for any $v \in \mathbb{R}^3$ with $v \neq 0$. (lines through the origin)
- 2-*dimensional subspaces*: $\operatorname{Span}\{u, v\}$ for any linearly independent $u, v \in \mathbb{R}^3$. (planes through the origin)
- 3-*dimensional subspaces*: Only $\mathbb{R}^3$ itself.

THEOREM 4.13. *Let V be an n-dimensional vector space, $n \geq 1$. Any linearly independent set of exactly n elements in V is automatically a basis for V. Any set of exactly n elements that spans V is automatically a basis for V.*

PROOF. By theorem 4.12 and theorem 4.5. □

COROLLARY. *If V is an n-dimensional vector space, then only n-dimensional subspace of V is itself.*

The Dimensions of Nul A, Col A, and Row A.

DEFINITION. *Let A be a matrix.*

- *The **rank** of A is*
$$\operatorname{rank} A = \dim \operatorname{Col} A.$$
- *The **nullity** of A is*
$$\operatorname{nullity} A = \dim \operatorname{Nul} A.$$

OBSERVATION.

- $\operatorname{rank} A$ *is the number of pivot columns of A.*
- $\operatorname{nullity} A$ *is the number of free variables of $Ax = 0$.*
- $\dim \operatorname{Row} A = \dim \operatorname{Col} A = \operatorname{rank} A$.

EXAMPLE. Let
$$A = \begin{bmatrix} 1 & -1 & -1 & 0 & 2 \\ 3 & -3 & -3 & 0 & 6 \\ 2 & 0 & 1 & 1 & 9 \\ -1 & 7 & 0 & 6 & -7 \end{bmatrix}$$
Then,
$$A \sim \begin{bmatrix} 1 & -1 & -1 & 0 & 2 \\ 0 & 2 & 3 & 1 & 5 \\ 0 & 0 & -10 & 3 & -20 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$
So $\operatorname{rank} A = 3$ and $\operatorname{nullity} A = 2$.

THEOREM 4.14 (Rank-nullity theorem). *Let A be an $m \times n$ matrix. Then*
$$\operatorname{rank} A + \operatorname{nullity} A = n.$$

PROOF. $\operatorname{rank} A$ is the number of pivot columns of A and $\operatorname{nullity} A$ is the number of free variables of $Ax = 0$ which is the number of non-pivot columns of A. Thus, $\operatorname{rank} A + \operatorname{nullity} A$ is the number of columns of A, which is n. □

EXAMPLE. If A is a 9×12 matrix with nullity 4, what is the rank of A?

SOLUTION. $12 - 4 = 8$.

EXAMPLE. Could a 7×12 matrix have nullity 4?

SOLUTION. No. If A is a 7×12 matrix with nullity 4, then $\operatorname{rank} A = 12 - 4 = 8$. But $\operatorname{Col} A$ is a subspace of $\mathbb{R}^7$. So, $\operatorname{rank} A \leq 7$.

EXAMPLE. Suppose you are given a homogeneous system of 30 equations in 32 variables and you have found two solutions that are not multiples, and all other solutions can be constructed by adding together appropriate multiples of these two solutions. Can you be certain that an associated nonhomogeneous system (with the same coefficients) has a solution?

SOLUTION. Yes. Let A be the coefficient matrix of the system. Then A is 30×32. The given information implies that the two solutions are linearly independent and span $\operatorname{Nul} A$. So $\operatorname{nullity} A = 2$ and $\operatorname{rank} A = 30$.
Since $\mathbb{R}^{30}$ is the only subspace of $\mathbb{R}^{30}$ whose dimension is 30, we have $\operatorname{Col} A = \mathbb{R}^{30}$. Then for any b, the nonhomogeneous equation $Ax = b$ is consistent.

Invertible Matrix Theorem, revisited. We are going to extend the invertible matrix theorem, theorem 2.8 (page 54 on the handwritten notes). We already stated (a)-(l), now we add (m)-(q).

THEOREM (The Invertible Matrix Theorem). *Let A be an $n \times n$ matrix. The following are equivalent.*

(a) *A is an invertible matrix.*
(b) *$A \sim I_n$.*
(c) *A has n pivot positions.*
(d) *The equation $Ax = 0$ has only the trivial solution.*
(e) *The columns of A form a linearly independent set.*
(f) *The linear transformation $x \mapsto Ax$ is one-to-one.*
(g) *The equation $Ax = b$ has at least one solution for each $b \in \mathbb{R}^n$.*
(h) *The columns of A span $\mathbb{R}^n$.*
(i) *The linear transformation $x \mapsto Ax$ is onto.*
(j) *There is an $n \times n$ matrix C such that $CA = I_n$.*
(k) *There is an $n \times n$ matrix D such that $AD = I_n$.*
(l) *A^T is invertible.*
(m) *The columns of A form a basis for $\mathbb{R}^n$.*
(n) $\operatorname{Col} A = \mathbb{R}^n$.
(o) $\operatorname{rank} A = n$.
(p) $\operatorname{nullity} A = 0$.
(q) $\operatorname{Nul} A = \{0\}$.

PROOF. We have already proved the equivalence of (a)-(b)-(c)-...-(l).
We showed (b) $\iff$ (m) after defining the basis.
(h) $\implies$ (n) by definition of $\operatorname{Col} A$.
(n) $\implies$ (o) by definition of $\operatorname{rank} A$.
(o) $\implies$ (p) by the rank-nullity theorem.
(p) $\implies$ (q) because $\{0\}$ is the only subspace with dimension 0.
(q) $\implies$ (d) by definition of $\operatorname{Nul} A$. □

Inverse of an isomorphism. If $T : V \to W$ is an isomorphism, then we can define its inverse. Since T is an isomorphism, for every $w \in W$, there is a unique $v \in V$ such that $T(v) = w$. Call $T^{-1}(w) := v$. Then

$$\begin{aligned} T^{-1} : W &\to V \\ w &\mapsto T^{-1}(w) \end{aligned}$$

is also an isomorphism.

EXAMPLE. If V is a vector space with basis $\mathcal{B} = \{b_1, \dots, b_n\}$, then the isomorphism

$$\begin{aligned} V &\longrightarrow \mathbb{R}^n \\ u &\longmapsto [u]_{\mathcal{B}} \end{aligned}$$

has the inverse

$$\begin{aligned} \mathbb{R}^n &\longrightarrow V \\ \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} &\longmapsto c_1 b_1 + \dots + c_n b_n. \end{aligned}$$

Rank-nullity Theorem for Linear Transformations. Now we generalize the rank-nullity theorem for linear transformations between vector spaces.

Let V be a vector space with basis

$$\mathcal{B} = \{b_1, \dots, b_n\}$$

and W be a vector space with basis

$$\mathcal{D} = \{d_1, \dots, d_m\}$$

and

$$T : V \longrightarrow W$$

be a linear transformation.

Consider the transformation

$$\widetilde{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

defined by

$$\widetilde{T}([v]_{\mathcal{B}}) = [T(v)]_{\mathcal{D}}.$$

So, we have the following picture.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, leftrightarrow] & W \arrow[d, leftrightarrow] \\ \mathbb{R}^n \arrow[r, "\widetilde{T}"] & \mathbb{R}^m \end{tikzcd} \qquad \begin{tikzcd} v \arrow[r, mapsto, "T"] \arrow[d, leftrightarrow] & T(v) \arrow[d, leftrightarrow] \\ {[v]_{\mathcal{B}}} \arrow[r, mapsto, "\widetilde{T}"] & {[T(v)]_{\mathcal{D}}} \end{tikzcd}$$

REMARK. Call

$$\begin{aligned} T_{\mathcal{B}} : V &\longrightarrow \mathbb{R}^n \\ v &\longmapsto [v]_{\mathcal{B}} \end{aligned}$$

and

$$\begin{aligned} T_{\mathcal{D}} : W &\longrightarrow \mathbb{R}^m \\ w &\longmapsto [w]_{\mathcal{D}} \end{aligned}$$

for the coordinate mappings determined by $\mathcal{B}$ and $\mathcal{D}$. Then $\widetilde{T}$ is the composition

$$T_{\mathcal{D}} \circ T \circ T_{\mathcal{B}}^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^m.$$

EXERCISE. Show the following.

- $\widetilde{T}$ is linear.
- $\dim \operatorname{Range}(T) = \dim \operatorname{Range}(\widetilde{T})$.
- $\dim \operatorname{Ker}(T) = \dim \operatorname{Ker}(\widetilde{T})$.

Let A be the standard matrix of the linear transformation $\widetilde{T}$. Then A is $m \times n$. Also,

$$\dim \operatorname{Range}(T) = \dim \operatorname{Range}(\widetilde{T}) = \dim \operatorname{Col} A = \operatorname{rank} A$$
$$\dim \operatorname{Ker}(T) = \dim \operatorname{Ker}(\widetilde{T}) = \dim \operatorname{Nul} A = \operatorname{nullity} A.$$

Hence, by the rank-nullity theorem for matrices,

$$\begin{aligned} \dim \operatorname{Range}(T) + \dim \operatorname{Ker}(T) &= \operatorname{rank} A + \operatorname{nullity} A \\ &= n \\ &= \dim V. \end{aligned}$$

We showed the following.

THEOREM (Rank-nullity theorem for linear transformations). *Let V and W be vector spaces, and $T : V \longrightarrow W$ be a linear transformation. Then*

$$\dim \operatorname{Range}(T) + \dim \operatorname{Ker}(T) = \dim V.$$

EXAMPLE. Let $T : \mathbb{P}_2 \longrightarrow \mathbb{R}^3$ be given by

$$T(p) = \begin{bmatrix} p(-1) \\ p(0) \\ p(1) \end{bmatrix}.$$

If $p = ax^2 + bx + c$, then $p(-1) = a - b + c$, $p(0) = c$, $p(1) = a + b + c$.

$$\begin{aligned} T(p) = 0 &\iff p(-1) = p(0) = p(1) = 0 \\ &\iff a = b = c = 0 \\ &\iff p = 0. \end{aligned}$$

Hence

$$\begin{aligned} \operatorname{Ker} T = \{0\} &\implies \dim \operatorname{Ker} T = 0 \\ &\implies \dim \operatorname{Range} T = \dim \mathbb{P}_2 - 0 = 3 \\ &\implies \operatorname{Range} T = \mathbb{R}^3. \end{aligned}$$

EXAMPLE. Let $T : \mathbb{P}_n \longrightarrow \mathbb{P}_{n-1}$ be given by

$$T(p) = \frac{dp}{dx}$$

that is

$$T(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n) = a_1 + 2a_2 x + 3a_3 x^2 + \ldots + n a_n x^{n-1}.$$

Then,

$$T(p) = 0 \iff p \in \mathbb{P}_0.$$

Hence

$$\begin{aligned} \operatorname{Ker} T = \mathbb{P}_0 &\implies \dim \operatorname{Ker} T = 1 \\ &\implies \dim \operatorname{Range} T = \dim \mathbb{P}_n - 1 = n \\ &\implies \operatorname{Range} T = \mathbb{P}_{n-1}. \end{aligned}$$

Isomorphic vector spaces.

DEFINITION. *The vector spaces V and W are called* ***isomorphic*** *if there is an isomorphism $V \longrightarrow W$. In this case, we denote*

$$V \cong W.$$

EXERCISE. Let $T_1 : V_1 \longrightarrow V_2$ and $T_2 : V_2 \longrightarrow V_3$ be isomorphisms. Show that the composition

$$T_2 \circ T_1 : V_1 \longrightarrow V_3$$

given by

$$(T_2 \circ T_1)(u) = T_2(T_1(u))$$

is also an isomorphism.

OBSERVATION. *Let V and W be finite-dimensional vector spaces.*

$$V \cong W \iff \dim V = \dim W.$$

PROOF. If $V \cong W$, then there is an isomorphism $T : V \longrightarrow W$. Since T is an isomorphism, $\operatorname{Ker} T = \{0\}$ and $\operatorname{Range} T = W$. Hence by rank-nullity theorem,

$$\dim W = \dim \operatorname{Range} T = \dim V - \dim \operatorname{Ker} T = \dim V - 0 = \dim V.$$

If $\dim V = \dim W = n$, then there is a basis $\mathcal{B}$ for V and a basis $\mathcal{D}$ for W such that $|\mathcal{B}| = |\mathcal{D}| = n$. Consider the isomorphisms

$$\begin{aligned} T_{\mathcal{B}} : V &\longrightarrow \mathbb{R}^n \\ v &\longmapsto [v]_{\mathcal{B}} \end{aligned}$$

and

$$\begin{aligned} T_{\mathcal{D}} : W &\longrightarrow \mathbb{R}^n \\ w &\longmapsto [w]_{\mathcal{D}}. \end{aligned}$$

Then

$$T_{\mathcal{D}}^{-1} \circ T_{\mathcal{B}} : V \longrightarrow W$$

is an isomorphism. $\square$

4.6. Change of Basis

We have seen in section 4.4, that given a basis $\mathcal{B}$ for an n-dimensional vector space V, the associated coordinate mapping onto $\mathbb{R}^n$ gives a coordinate system for V. Now, we will see that given two bases for V, how their coordinate systems are related.

EXAMPLE. Let $\mathcal{B} = \{b_1, b_2\}$ and $\mathcal{D} = \{d_1, d_2\}$ be two bases for a vector space V, such that

$$b_1 = 2d_1 - d_2 \quad \text{and} \quad b_2 = -3d_1 + 4d_2.$$

Let $u \in V$. Suppose $[u]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Find $[u]_{\mathcal{D}}$.

SOLUTION. Since $[u]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, we have $u = 2b_1 + 3b_2$. Then

$$\begin{aligned} [u]_{\mathcal{D}} &= [2b_1 + 3b_2]_{\mathcal{D}} \\ &= 2[b_1]_{\mathcal{D}} + 3[b_2]_{\mathcal{D}} \\ &= 2\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 3\begin{bmatrix} -3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 10 \end{bmatrix}. \end{aligned}$$

Notice that we had

$$\begin{aligned} [u]_{\mathcal{D}} &= 2[b_1]_{\mathcal{D}} + 3[b_2]_{\mathcal{D}} \\ &= \begin{bmatrix} [b_1]_{\mathcal{D}} & [b_2]_{\mathcal{D}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} [b_1]_{\mathcal{D}} & [b_2]_{\mathcal{D}} \end{bmatrix} [u]_{\mathcal{B}}. \end{aligned}$$

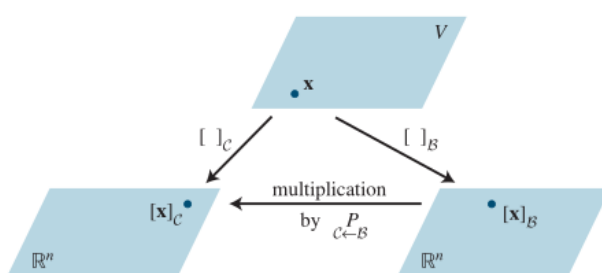
THEOREM 4.15. *Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{D} = \{d_1, \dots, d_n\}$ be bases for a vector space V. Then there is a unique $n \times n$ matrix $P_{\mathcal{D} \leftarrow \mathcal{B}}$ such that*

$$[u]_{\mathcal{D}} = P_{\mathcal{D} \leftarrow \mathcal{B}}[u]_{\mathcal{B}}$$

for all $u \in V$. This matrix is given by

$$P_{\mathcal{D} \leftarrow \mathcal{B}} = \begin{bmatrix} [b_1]_{\mathcal{D}} & [b_2]_{\mathcal{D}} & \dots & [b_n]_{\mathcal{D}} \end{bmatrix}.$$

We call $P_{\mathcal{D} \leftarrow \mathcal{B}}$ the *change-of-coordinates matrix* from $\mathcal{B}$ to $\mathcal{D}$.



Proof. Consider the transformation

$$\begin{aligned} \widetilde{T} : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ [u]_{\mathcal{B}} &\longmapsto [u]_{\mathcal{D}}. \end{aligned}$$

Call

$$\begin{aligned} T_{\mathcal{B}} : V &\longrightarrow \mathbb{R}^n \\ u &\longmapsto [u]_{\mathcal{B}} \end{aligned}$$

and

$$\begin{aligned} T_{\mathcal{D}} : V &\longrightarrow \mathbb{R}^n \\ u &\longmapsto [u]_{\mathcal{D}} \end{aligned}$$

for the coordinate mappings determined by $\mathcal{B}$ and $\mathcal{D}$. Then $\widetilde{T}$ is the composition

$$T_{\mathcal{D}} \circ T_{\mathcal{B}}^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

since $[u]_{\mathcal{B}} \xmapsto{T_{\mathcal{B}}^{-1}} u \xmapsto{T_{\mathcal{D}}} [u]_{\mathcal{D}}$. Then, $\widetilde{T}$ is an isomorphism. Moreover, its standard matrix is

$$P_{\mathcal{D}\leftarrow\mathcal{B}} = \begin{bmatrix} [b_1]_{\mathcal{D}} & [b_2]_{\mathcal{D}} & \dots & [b_n]_{\mathcal{D}} \end{bmatrix}$$

because

$$\widetilde{T}(e_i) = T_{\mathcal{D}}(T_{\mathcal{B}}^{-1}(e_i)) = T_{\mathcal{D}}(b_i) = [b_i]_{\mathcal{D}}.$$

□

Example. Consider the bases $\mathcal{B} = \{1 - 2x + x^2, -1 + 3x - x^2, 4x - x^2\}$ and $\mathcal{D} = \{1, 3x, 5x^2\}$ for $V = \mathbb{P}_2$. Then

$$[1 - 2x + x^2]_{\mathcal{D}} = \begin{bmatrix} 1 \\ -2/3 \\ 1/5 \end{bmatrix}$$

$$[-1 + 3x - x^2]_{\mathcal{D}} = \begin{bmatrix} -1 \\ 1 \\ -1/5 \end{bmatrix}$$

$$[4x - x^2]_{\mathcal{D}} = \begin{bmatrix} 0 \\ 4/3 \\ -1/5 \end{bmatrix}$$

so

$$P_{\mathcal{D}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & -1 & 0 \\ -2/3 & 1 & 4/3 \\ 1/5 & -1/5 & -1/5 \end{bmatrix}.$$

- Since $\{b_1, \dots, b_n\}$ is linearly independent, so is $\{[b_1]_{\mathcal{D}}, \dots, [b_n]_{\mathcal{D}}\}$. Then the square matrix $P_{\mathcal{D}\leftarrow\mathcal{B}}$ is invertible.

Multiplying $[u]_{\mathcal{D}} = P_{\mathcal{D}\leftarrow\mathcal{B}}[u]_{\mathcal{B}}$ with $P_{\mathcal{D}\leftarrow\mathcal{B}}^{-1}$ we have

$$P_{\mathcal{D}\leftarrow\mathcal{B}}^{-1}[u]_{\mathcal{D}} = [u]_{\mathcal{B}}.$$

This shows the following.

OBSERVATION. $P^{-1}_{\mathscr{D}\leftarrow\mathscr{B}} = P_{\mathscr{B}\leftarrow\mathscr{D}}$.

OBSERVATION. *Let $\mathscr{B}$, $\mathscr{C}$ and $\mathscr{D}$ be bases for a finite-dimensional vector space V. Then*

$$P_{\mathscr{D}\leftarrow\mathscr{C}}P_{\mathscr{C}\leftarrow\mathscr{B}} = P_{\mathscr{D}\leftarrow\mathscr{B}}.$$

PROOF. $(P_{\mathscr{D}\leftarrow\mathscr{C}}P_{\mathscr{C}\leftarrow\mathscr{B}})e_i = P_{\mathscr{D}\leftarrow\mathscr{C}}[b_i]_{\mathscr{C}} = [b_i]_{\mathscr{D}}$ for all $1 \leq i \leq n$. $\square$

Change of Basis in $\mathbb{R}^n$. If we take $V = \mathbb{R}^n$ and ε to be the standard basis $\{e_1, \ldots, e_n\}$ then $P_{\varepsilon\leftarrow\mathscr{B}} = P_{\mathscr{B}}$.

We will look at the coordinate change between two non-standard bases of $\mathbb{R}^n$.

EXAMPLE. Let $b_1 = \begin{bmatrix} -1 \\ -3 \end{bmatrix}$, $b_2 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$, $d_1 = \begin{bmatrix} 1 \\ -7 \end{bmatrix}$, $d_2 = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$. Then $\mathscr{B} = \{b_1, b_2\}$ and $\mathscr{D} = \{d_1, d_2\}$ are two bases for $\mathbb{R}^2$. Find $P_{\mathscr{D}\leftarrow\mathscr{B}}$.

SOLUTION. We need to find $[b_1]_{\mathscr{D}}$ and $[b_2]_{\mathscr{D}}$. For this we will solve the following systems.

$$\begin{aligned} x_1 d_1 + x_2 d_2 &= b_1 \\ y_2 d_1 + y_2 d_2 &= b_2 \end{aligned}$$

They correspond to the augmented matrices

$$\left[\begin{array}{cc|c} d_1 & d_2 & b_1 \end{array}\right] \text{ and } \left[\begin{array}{cc|c} d_1 & d_2 & b_2 \end{array}\right].$$

We can reduce these two simultaneously, by viewing this as the following augmented matrix.

$$\left[\begin{array}{cc|cc} d_1 & d_2 & b_1 & b_2 \end{array}\right].$$

Then,

$$\begin{aligned} \left[\begin{array}{cc|cc} d_1 & d_2 & b_1 & b_2 \end{array}\right] &= \left[\begin{array}{cc|cc} 1 & -2 & -1 & -3 \\ -7 & 7 & -3 & 5 \end{array}\right] \\ &\sim \left[\begin{array}{cc|cc} 1 & 0 & 13/7 & 11/7 \\ 0 & 1 & 10/7 & 16/7 \end{array}\right] \end{aligned}$$

Thus

$$[b_1]_{\mathscr{D}} = \begin{bmatrix} 13/7 \\ 10/7 \end{bmatrix} \text{ and } [b_2]_{\mathscr{D}} = \begin{bmatrix} 11/7 \\ 16/7 \end{bmatrix}.$$

Hence

$$P_{\mathscr{D}\leftarrow\mathscr{B}} = \begin{bmatrix} [b_1]_{\mathscr{D}} & [b_2]_{\mathscr{D}} \end{bmatrix} = \begin{bmatrix} 13/7 & 11/7 \\ 10/7 & 16/7 \end{bmatrix}.$$

We can generalize this solution technique as follows.

OBSERVATION. *If $\mathscr{B} = \{b_1, \ldots, b_n\}$ and $\mathscr{D} = \{d_1, \ldots, d_n\}$ are two bases for $\mathbb{R}^n$, then*

$$\left[\begin{array}{ccc|ccc} d_1 & \ldots & d_n & b_1 & \ldots & b_n \end{array}\right] \sim \left[\begin{array}{c|c} I_n & P_{\mathscr{D}\leftarrow\mathscr{B}} \end{array}\right]$$

EXAMPLE. Let $b_1 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$, $b_2 = \begin{bmatrix} 5 \\ -2 \\ 6 \end{bmatrix}$, $b_3 = \begin{bmatrix} 7 \\ 9 \\ 7 \end{bmatrix}$, $d_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $d_2 = \begin{bmatrix} 2 \\ 5 \\ -2 \end{bmatrix}$, $d_3 = \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}$. Then $\mathscr{B} = \{b_1, b_2, b_3\}$ and $\mathscr{D} = \{d_1, d_2, d_3\}$ are bases for $\mathbb{R}^3$. Find $P_{\mathscr{D}\leftarrow\mathscr{B}}$.

SOLUTION.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -3 & 5 & 7 \\ 2 & 5 & -3 & 0 & -2 & 9 \\ -1 & -2 & 1 & 1 & 6 & 7 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -3 & -37 & -67 \\ 0 & 1 & 0 & 0 & 21 & 37 \\ 0 & 0 & 1 & -2 & 11 & 14 \end{array}\right].$$

Thus

$$P_{\mathscr{D}\leftarrow\mathscr{B}} = \begin{bmatrix} -3 & -37 & -67 \\ 0 & 21 & 37 \\ -2 & 11 & 14 \end{bmatrix}.$$

OBSERVATION. *Let $\mathscr{B}$ and $\mathscr{D}$ be bases for $\mathbb{R}^n$. Then*

$$P_{\mathscr{D}\leftarrow\mathscr{B}} = P_{\mathscr{D}}^{-1} P_{\mathscr{B}}.$$

PROOF. Let $\varepsilon = \{e_1, \ldots, e_n\}$. Then

$$P_{\mathscr{B}} = P_{\varepsilon\leftarrow\mathscr{B}} = P_{\varepsilon\leftarrow\mathscr{D}} P_{\mathscr{D}\leftarrow\mathscr{B}} = P_{\mathscr{D}} P_{\mathscr{D}\leftarrow\mathscr{B}}.$$

$\square$

CHAPTER 3

Determinants

3.1. Introduction to Determinants

Recall from Theorem 4 of Section 2.2 that a 2×2 matrix is invertible if and only if its determinant is nonzero. To extend this useful fact to larger matrices, we need a definition for the determinant of an $n \times n$ matrix.

Consider $n \times n$ matrix $A = [a_{ij}]$ with $a_{11} \neq 0$.

n=2 case. Recall that if $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\det A = a_{11}a_{22} - a_{12}a_{21}$. Also we showed that A is invertible if and only if $\det A \neq 0$. An alternative proof of this is the following.

$$\begin{aligned} A &\sim \begin{bmatrix} a_{11} & a_{12} \\ a_{11}a_{21} & a_{11}a_{22} \end{bmatrix} \\ &\sim \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix} = \widetilde{A}. \end{aligned}$$

(Recall that we assumed $a_{11} \neq 0$.)

Hence

$$\begin{aligned} A \text{ is invertible} &\iff \widetilde{A} \text{ is invertible} \\ &\iff a_{11}a_{22} - a_{12}a_{21} \neq 0. \end{aligned}$$

n=3 case. We have $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, $a_{11} \neq 0$.

$$\begin{aligned} A &\sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{bmatrix} \\ &\sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & a_{11}a_{32} - a_{12}a_{31} & a_{11}a_{33} - a_{13}a_{31} \end{bmatrix} = \widetilde{A}. \end{aligned}$$

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

Hence

$$\begin{aligned} A \text{ is invertible} &\iff \widetilde{A} \text{ is invertible} \\ &\iff \widetilde{A} \text{ has 3 pivots} \\ &\iff B \text{ has 2 pivots} \\ &\iff B \text{ is invertible} \\ &\iff \det B \neq 0. \end{aligned}$$

We have

$$\begin{aligned} \det B =&(a_{11}a_{22} - a_{12}a_{21})(a_{11}a_{33} - a_{13}a_{31}) - (a_{11}a_{23} - a_{13}a_{21})(a_{11}a_{32} - a_{12}a_{31}) \\ =& + a_{11}^2 a_{22}a_{33} - a_{11}a_{13}a_{22}a_{31} - a_{11}a_{12}a_{21}a_{33} + a_{12}a_{13}a_{21}a_{31} \\ & - a_{11}^2 a_{23}a_{32} + a_{11}a_{12}a_{23}a_{31} + a_{11}a_{13}a_{21}a_{32} - a_{12}a_{13}a_{21}a_{31} \\ =& a_{11}^2(a_{22}a_{33} - a_{23}a_{32}) - a_{11}a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{11}a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ =& a_{11}(a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})). \end{aligned}$$

So

$$\begin{aligned} \frac{\det B}{a_{11}} &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}. \end{aligned}$$

Call

$$A_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, A_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}, A_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

Thus

$$\begin{aligned} A \text{ is invertible} &\iff \det B \neq 0 \\ &\iff a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \neq 0. \end{aligned}$$

We define

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}.$$

General case.

DEFINITION. *For an $n \times n$ matrix A, we define A_{ij} to be the $(n-1) \times (n-1)$ matrix formed by deleting the i-th row and j-th column of A.*

For instance, if

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

then

$$A_{32} = \begin{bmatrix} 1 & 5 & -7 \\ 0 & 4 & -8 \\ 6 & 0 & -3 \end{bmatrix}.$$

DEFINITION.

- *For $n \geq 2$, the* ***determinant*** *of an $n \times n$ matrix $A = [a_{ij}]$ is defined as*

$$\begin{aligned}\det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \ldots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}.\end{aligned}$$

- *The* ***determinant*** *of a 1×1 matrix is defined as* $\det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11}$.

Example.

$$\begin{aligned}\det \begin{bmatrix} 1 & 5 & 0 \\ 2 & 0 & -1 \\ 3 & 6 & -1 \end{bmatrix} &= 1 \det \begin{bmatrix} 0 & -1 \\ 6 & -1 \end{bmatrix} - 5 \det \begin{bmatrix} 2 & -1 \\ 3 & -1 \end{bmatrix} + 0 \det \begin{bmatrix} 2 & 0 \\ 3 & 6 \end{bmatrix} \\ &= 1 \cdot 6 - 5 \cdot 1 \\ &= 1.\end{aligned}$$

Definition. *For an $n \times n$ matrix A, we define the (i, j)-**cofactor** of A to be*

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

With this notation, if $A = [a_{ij}]$,

$$\begin{aligned}\det A &= a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n} \\ &= \sum_{j=1}^{n} a_{1j}C_{1j}.\end{aligned}$$

This formula is called a *cofactor expansion* across the first row of A. The following theorem says that the determined can be computed by a cofactor expansion across any row or down any column.

Theorem 3.1. *Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then,*

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$$

and

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$$

for any i and j.

We omit the proof in our course. Curious reader may read the proof in the book Linear algebra by Friedberg, Insel, Spence (4th edition), section 4.2, 4.3.

Example. Let $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -3 & 0 \end{bmatrix}$. By definition,

$$\begin{aligned}\det A &= 1 \det \begin{bmatrix} 4 & -1 \\ -3 & 0 \end{bmatrix} - 5 \det \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} + 0 \det \begin{bmatrix} 2 & 4 \\ 0 & -3 \end{bmatrix} \\ &= 1 \cdot (-3) - 5 \cdot 0 \\ &= -3.\end{aligned}$$

Using the cofactor expansion across the third row,

$$\begin{aligned}\det A &= 0 \cdot C_{31} + (-3) \cdot C_{32} + 0 \cdot C_{33} \\ &= (-3) \cdot C_{32} \\ &= (-3)(-1)^5 \det \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \\ &= (-3)(-1)(-1) \\ &= -3.\end{aligned}$$

Definition. *Let A be a square matrix.*

- *A is called* ***upper triangular*** *if all entries below the main diagonal are all 0.*
- *A is called* ***lower triangular*** *if all entries above the main diagonal are all 0.*
- *A is called* ***triangular*** *if it is upper triangular or lower triangular.*

Example.

$$A = \begin{bmatrix} 2 & 5 & -7 & 3 & 6 \\ 0 & -3 & 4 & -1 & 9 \\ 0 & 0 & 3 & 7 & -8 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & -4 \end{bmatrix}$$

is a (upper) triangular matrix.

Using the cofactor expansion across the first column,

$$\begin{aligned}\det A &= 2 \det \begin{bmatrix} -3 & 4 & -1 & 9 \\ 0 & 3 & 7 & -8 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & -4 \end{bmatrix} \\ &= 2 \cdot (-3) \det \begin{bmatrix} 3 & 7 & -8 \\ 0 & -1 & 2 \\ 0 & 0 & -4 \end{bmatrix} \\ &= 2 \cdot (-3) \cdot 3 \det \begin{bmatrix} -1 & 2 \\ 0 & -4 \end{bmatrix} \\ &= 2 \cdot (-3) \cdot 3 \cdot (-1) \cdot (-4) \\ &= 72.\end{aligned}$$

We can generalize this as follows.

Theorem 3.2. *Determinant of a (square) triangular matrix, is the product of the entries on the main diagonal.*

Corollary. $\det I_n = 1$.

3.2. Properties of Determinants

OBSERVATION. *Let A be an $n \times n$ matrix, with $n \geq 2$. If A has two identical rows, then $\det A = 0$.*

PROOF. We will prove this by induction.

If $n = 2$, then A is of the form $\begin{bmatrix} a & b \\ a & b \end{bmatrix}$. Then $\det A = ab - ab = 0$.

Assume $n \geq 3$ and assume that the statement is true for $(n-1) \times (n-1)$ matrices. Let A be an $n \times n$ matrix, with $\text{row}_r A = \text{row}_s A$. Let $i \in \{1, \dots, n\} \setminus \{r, s\}$. Then each A_{ij} is an $(n-1) \times (n-1)$ matrix having two identical rows. Then $\det A_{ij} = 0$. Then, cofactor expansion across the i-th row of A gives $\det A = 0$. $\square$

THEOREM 3.3. *Let A be a square matrix.*

(i) If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$.
(ii) If two rows of A are interchanged to produce B, then $\det B = -\det A$.
(iii) If one row of A is multiplied by c to produce B, then $\det B = c \det A$.

PROOF. (i) Say $A = \begin{bmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix}$ and $B = \begin{bmatrix} r_1 \\ \vdots \\ r_i + cr_j \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix}$. Looking at cofactor expansion across the i-th row of B,

$$\det B = \det \begin{bmatrix} r_1 \\ \vdots \\ r_i + cr_j \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix} = \det \begin{bmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix} + c \det \begin{bmatrix} r_1 \\ \vdots \\ r_j \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix}$$
$$= \det A + c \cdot 0 = \det A.$$

(ii) Say $A = \begin{bmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_j \\ \vdots \\ r_n \end{bmatrix}$ and $B = \begin{bmatrix} r_1 \\ \vdots \\ r_j \\ \vdots \\ r_i \\ \vdots \\ r_n \end{bmatrix}$. Looking at cofactor expansion across the i-th row of A, and j-th row of B, we have $\det B = -\det A$.

(iii) Say $A = \begin{bmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_n \end{bmatrix}$ and $B = \begin{bmatrix} r_1 \\ \vdots \\ cr_i \\ \vdots \\ r_n \end{bmatrix}$. Looking at cofactor expansion across the i-th row of A and B, we have $\det B = c \det A$.

□

This theorem gives an efficient way to calculate determinants.

Example.

$$\begin{aligned}
\det \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix} &= \det \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ 0 & 3 & 2 \end{bmatrix} \\
&= \det \begin{bmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{bmatrix} \\
&= -\det \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{bmatrix} \\
&= (-1) \cdot 1 \cdot 3 \cdot (-5) = 15.
\end{aligned}$$

Let A be an $n \times n$ matrix. It can be reduced to an echelon form U by row replacements and row interchanges, using the row reduction algorithm. If there are r interchanges, then

$$\det A = (-1)^r \det U.$$

Since U is in echelon form, it is triangular, so $\det U$ is the product of its diagonal entries. Therefore,

$$\begin{aligned}
A \text{ is invertible} &\iff U \text{ is invertible} \\
&\iff \text{all pivots of } U \text{ are non-zero} \\
&\iff \text{all diagonal entries of } U \text{ are non-zero} \\
&\iff \det U \neq 0 \\
&\iff \det A \neq 0.
\end{aligned}$$

We showed the following.

THEOREM 3.4. *A square matrix A is invertible if and only if* $\det A \neq 0$.

EXAMPLE. Let $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$. Then,

$$\det A = \det \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix} = 0.$$

So A is not invertible.

The calculations in the next example combine the power of row operations with the strategy of using zero entries in cofactor expansions.

EXAMPLE.

$$\begin{aligned}
\det \begin{bmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{bmatrix} &= \det \begin{bmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{bmatrix} \\
&= -2 \det \begin{bmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{bmatrix} \\
&= -2 \det \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 5 \\ 0 & -3 & 1 \end{bmatrix} \\
&= 2 \det \begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 1 \\ 0 & 0 & 5 \end{bmatrix} \\
&= 2 \cdot 1 \cdot (-3) \cdot 5 \\
&= -30.
\end{aligned}$$

Column Operations. By theorem 3.1, the determined can be computed by a cofactor expansion using rows or columns. This implies the following.

THEOREM 3.5. *If A is a square matrix, then* $\det A^T = \det A$.

This theorem implies that, to find determinant, instead of row operations we can use column operations as well.

EXAMPLE.

$$\det\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 5 & 0 \\ 0 & 3 & 6 & 1 \\ -1 & -2 & -6 & 0 \end{bmatrix} = \det\begin{bmatrix} 1 & 0 & 3 & 4 \\ 1 & 0 & 5 & 0 \\ 0 & 3 & 6 & 1 \\ -1 & 0 & -6 & 0 \end{bmatrix}$$
$$= -3\det\begin{bmatrix} 1 & 3 & 4 \\ 1 & 5 & 0 \\ -1 & -6 & 0 \end{bmatrix}$$
$$= (-3)\cdot 4\det\begin{bmatrix} 1 & 5 \\ -1 & -6 \end{bmatrix}$$
$$= (-3)\cdot 4\cdot(-6+5)$$
$$= 12.$$

Determinants and elementary matrices. Theorem 3.3 can be reformulated as follows.

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det EA = (\det E)(\det A)$$

and

$$\det E = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ c & \text{if } E \text{ is a scale by } c. \end{cases}$$

Determinants and Matrix Products.

OBSERVATION. *Let A and B be $n \times n$ matrices. Then AB is invertible if and only if A and B are both invertible.*

PROOF. If A and B are invertible, then by theorem 2.6(ii), AB is invertible.

If AB is invertible, then there is an $n \times n$ matrix C such that $(AB)C = I_n = C(AB)$. Then BC is the inverse of A and CA is the inverse of B. $\square$

THEOREM 3.6. *If A and B are $n \times n$ matrices, then*

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

PROOF. If at least one of A and B is not invertible, then AB is not invertible. In this case, $\det AB = 0 = \det A \det B$.

Suppose both A and B are invertible. Since A is invertible,

$$A = E_1 E_2 \ldots E_m$$

for some elementary matrices $E_1, E_2, \ldots, E_m$. Then,

$$\begin{aligned}
\det(AB) &= \det(E_1E_2 \ldots E_mB) \\
&= \det E_1 \det(E_2 \ldots E_mB) \\
&= \det E_1 \det E_2 \det(E_3 \ldots E_mB) \\
&\ldots \\
&= \det E_1 \det E_2 \det E_3 \ldots \det E_m \det B \\
&= \det E_1 \det E_2 \ldots \det(E_{m-1}E_m) \det B \\
&\ldots \\
&= \det(E_1E_2 \ldots E_m) \det B \\
&= \det A \det B.
\end{aligned}$$

□

EXAMPLE. Let $A = \begin{bmatrix} 1 & 0 \\ -1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ -7 & -8 \end{bmatrix}$. Then $AB = \begin{bmatrix} 2 & 1 \\ -23 & -25 \end{bmatrix}$. We have $\det A = 3$, $\det B = -9$ and $\det AB = -27$.

COROLLARY. *If A is an invertible square matrix, then*

$$\det(A^{-1}) = \frac{1}{\det A}.$$

PROOF. Since $AA^{-1} = I$, we have $\det A \cdot \det A^{-1} = \det I = 1$. □

3.3. Cramer's Rule

For an $n \times n$ matrix $A = \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix}$ and a vector $b \in \mathbb{R}^n$, we define

$$A_i(b) = \begin{bmatrix} a_1 & \ldots & a_{i-1} & b & a_{i+1} & \ldots & a_n \end{bmatrix}.$$

THEOREM 3.7 (Cramer's Rule). *Let A be an invertible $n \times n$ matrix. For any b in $\mathbb{R}^n$, the unique solution $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ of $Ax = b$ is given by*

$$x_i = \frac{\det A_i(b)}{\det A}.$$

EXAMPLE. Use Cramer's rule to solve the system

$$\begin{aligned}
3x_1 - 2x_2 &= 4 \\
-4x_1 + 5x_2 &= 18.
\end{aligned}$$

SOLUTION. View the system as $Ax = b$, where $A = \begin{bmatrix} 3 & -2 \\ -4 & 5 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 18 \end{bmatrix}$. Then $\det A = 15 - 8 = 7 \neq 0$.

We have $A_1(b) = \begin{bmatrix} 4 & -2 \\ 18 & 5 \end{bmatrix}$, $A_2(b) = \begin{bmatrix} 3 & 4 \\ -4 & 18 \end{bmatrix}$. Therefore

$$\begin{aligned} x_1 &= \frac{\det A_1(b)}{\det A} = \frac{20+36}{7} = \frac{56}{7} = 8 \\ x_2 &= \frac{\det A_2(b)}{\det A} = \frac{54+16}{7} = \frac{70}{7} = 10. \end{aligned}$$

Proof of Cramer's Rule. If $Ax = b$, then

$$\begin{aligned} A \cdot I_i(x) &= A \begin{bmatrix} e_1 & \dots & e_{i-1} & x & e_{i+1} & \dots & e_n \end{bmatrix} \\ &= \begin{bmatrix} Ae_1 & \dots & Ae_{i-1} & Ax & Ae_{i+1} & \dots & Ae_n \end{bmatrix} \\ &= \begin{bmatrix} a_1 & \dots & a_{i-1} & b & a_{i+1} & \dots & a_n \end{bmatrix} \\ &= A_i(b). \end{aligned}$$

Thus,

$$\det A \cdot \det I_i(x) = \det A_i(b).$$

Calculating $\det I_i(x)$ across i-th row, we deduce

$$\det I_i(x) = \det \begin{bmatrix} e_1 & \dots & e_{i-1} & x & e_{i+1} & \dots & e_n \end{bmatrix} = x_i.$$

Since A is invertible, $\det A \neq 0$. Hence

$$x_i = \det I_i(x) = \frac{\det A_i(b)}{\det A}.$$

□

Example. Consider the following system in which s is an unspecified parameter. Determine the values of s for which the system has a unique solution, and use Cramer's rule to describe the solution

$$\begin{aligned} 2sx_1 - 3x_2 &= 4 \\ -6x_1 + sx_2 &= 1. \end{aligned}$$

Solution. View the system as $Ax = b$, where $A = \begin{bmatrix} 2s & -3 \\ -6 & s \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$. Then $\det A = 2s^2 - 18 = 2(s-3)(s+3)$. So the system has a unique solution iff $s \neq \pm 3$.

We have $A_1(b) = \begin{bmatrix} 4 & -3 \\ 1 & s \end{bmatrix}$, $A_2(b) = \begin{bmatrix} 2s & 4 \\ -6 & 1 \end{bmatrix}$. Therefore

$$\begin{aligned} x_1 &= \frac{\det A_1(b)}{\det A} = \frac{4s+3}{2(s-3)(s+3)} \\ x_2 &= \frac{\det A_2(b)}{\det A} = \frac{2s+24}{2(s-3)(s+3)} = \frac{s+12}{(s-3)(s+3)}. \end{aligned}$$

A formula for A^{-1}. Let A be an invertible $n \times n$ matrix. The j-th column of A^{-1} is a vector x such that

$$Ax = e_j.$$

Then, by Cramer's rule, the (i,j)-entry of A^{-1} is

$$x_i = \frac{\det A_i(e_j)}{\det A} = \frac{(-1)^{i+j}\det A_{ji}}{\det A} = \frac{C_{ji}}{\det A}.$$

Hence

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & C_{31} & \dots & C_{n1} \\ C_{12} & C_{22} & C_{32} & \dots & C_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & C_{3n} & \dots & C_{nn} \end{bmatrix}.$$

Definition. *The* ***adjugate*** *(or* ***classical adjoint****) of an $n \times n$ matrix A is*

$$\operatorname{adj} A = \begin{bmatrix} C_{11} & C_{21} & C_{31} & \dots & C_{n1} \\ C_{12} & C_{22} & C_{32} & \dots & C_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & C_{3n} & \dots & C_{nn} \end{bmatrix}.$$

Theorem 3.8. *If A is an invertible matrix, then*

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A.$$

This also implies

$$A \cdot \operatorname{adj} A = \det A \cdot I.$$

Example. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 5 & -3 \\ 7 & -1 & 0 \end{bmatrix}$.

Then

$$C_{11} = \det\begin{bmatrix} 5 & -3 \\ -1 & 0 \end{bmatrix} = -3, \quad C_{12} = -\det\begin{bmatrix} -2 & -3 \\ 7 & 0 \end{bmatrix} = -21, \quad C_{13} = \det\begin{bmatrix} -2 & 5 \\ 7 & -1 \end{bmatrix} = -33,$$

$$C_{21} = -\det\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} = -3, \quad C_{22} = \det\begin{bmatrix} 1 & 3 \\ 7 & 0 \end{bmatrix} = -21, \quad C_{23} = -\det\begin{bmatrix} 1 & 2 \\ 7 & -1 \end{bmatrix} = 15,$$

$$C_{31} = \det\begin{bmatrix} 2 & 3 \\ 5 & -3 \end{bmatrix} = -21, \quad C_{32} = -\det\begin{bmatrix} 1 & 3 \\ -2 & -3 \end{bmatrix} = -3, \quad C_{33} = \det\begin{bmatrix} 1 & 2 \\ -2 & 5 \end{bmatrix} = 9.$$

So,

$$\operatorname{adj} A = \begin{bmatrix} -3 & -3 & -21 \\ -21 & -21 & -3 \\ -33 & 15 & 9 \end{bmatrix}.$$

Then

$$A \cdot \operatorname{adj} A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 5 & -3 \\ 7 & -1 & 0 \end{bmatrix} \begin{bmatrix} -3 & -3 & -21 \\ -21 & -21 & -3 \\ -33 & 15 & 9 \end{bmatrix} = \begin{bmatrix} -144 & 0 & 0 \\ 0 & -144 & 0 \\ 0 & 0 & -144 \end{bmatrix}.$$

Hence $\det A = -144$ and

$$A^{-1} = \frac{1}{-144} \begin{bmatrix} -3 & -3 & -21 \\ -21 & -21 & -3 \\ -33 & 15 & 9 \end{bmatrix} = \begin{bmatrix} 1/48 & 1/48 & 7/48 \\ 7/48 & 7/48 & 1/48 \\ 11/48 & -5/48 & -1/16 \end{bmatrix}.$$

Determinants as Area or Volume.

THEOREM 3.9. *If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$. In general, if A is an $n \times n$ matrix, the n-dimensional hypervolume of the n-dimensional parallelotope determined by the columns of A is $|\det A|$.*

HOMEWORK. Read the part "Determinants as Area or Volume" in Section 3.3 from the textbook.

CHAPTER 5

Eigenvalues and Eigenvectors

5.1. Eigenvectors and Eigenvalues

EXAMPLE. Let $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ and $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then

$$Au = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 5u.$$

DEFINITION. *Let A be an $n \times n$ matrix.*

- *An* ***eigenvector*** *of A is a nonzero vector x such that $Ax = \lambda x$ for some scalar λ.*
- *A scalar λ is called an* ***eigenvalue*** *of A if there is a nonzero vector x such that $Ax = \lambda x$; such an x is called an* ***eigenvector corresponding*** *to λ.*

EXAMPLE. Let $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ and $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as in the example above. Then

$$Au = 5u$$

so u is an eigenvector corresponding to an eigenvalue 5.

Let $v = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$. Then

$$Av = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -5 \\ -10 \end{bmatrix} \neq \lambda \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

for any λ, so v is not an eigenvector of A.

EXAMPLE. Show that 7 is an eigenvalue of the matrix $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ and find the corresponding eigenvectors.

SOLUTION. 7 is an eigenvalue of A if and only if the equation

$$Ax = 7x$$

has a non-trivial solution. This is equivalent to $Ax - 7x = 0$ which is

$$(A - 7I)x = 0.$$

To solve this homogeneous equation, form the matrix

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}.$$

The columns of $A-7I$ are linearly dependent, so $(A-7I)x = 0$ has nontrivial solutions. Thus 7 is an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$A - 7I = \left[\begin{array}{cc|c} -6 & 6 & 0 \\ 5 & -5 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

The general solution has the form $x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$.

The idea here applies to any square matrix. λ is an eigenvalue of a square matrix A if and only if $Ax = \lambda x$ has a non-trivial solution, that is $(A - \lambda I)x = Ax - \lambda x = 0$ has a non-trivial solution.

Observation. *λ is an eigenvalue of a square matrix A if and only if the equation*

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

has a non-trivial solution.

Eigenspace. We call eigenspace for the solution set of $(A - \lambda I)x = 0$.

Definition. *Let A be a square matrix and λ be an eigenvalue of A. The null space of the matrix $A - \lambda I$ is called the* ***eigenspace*** *of A corresponding to λ.*

Note that if A is $n \times n$, the eigenspace is a subspace of $\mathbb{R}^n$.

The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

$$\text{eigenspace of } A \text{ corresponding to } \lambda = \{0\} \cup \{\text{eigenvectors of } A \text{ corresponding to } \lambda\}$$

Example. In the example above, where $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ and $\lambda = 7$, the eigenspace of A corresponding to λ is

$$\left\{ x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} : x_2 \in \mathbb{R} \right\} = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Example. Let $A = \begin{bmatrix} 3 & 2 & 3 \\ 1 & 4 & 3 \\ 1 & 2 & 5 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

Solution. We need to solve $(A - 2I)x = 0$.

$$\left[\begin{array}{c|c} A - 2I & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The solutions are given by $x_1 + 2x_2 + 3x_3 = 0$, that is

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix},$$

so the corresponding eigenspace is

$$\left\{ x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} : x_2, x_3 \in \mathbb{R} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thus

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for the corresponding eigenspace.

5.2. The Characteristic Equation

Let A be a square matrix. We can continue our reasoning as follows.

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff Ax = \lambda x \text{ has a non-trivial solution} \\ &\iff (A - \lambda I)x = 0 \text{ has a non-trivial solution} \\ &\iff A - \lambda I \text{ is not invertible} \\ &\iff \det(A - \lambda I) = 0. \end{aligned}$$

Definition. *For a square matrix A the equation $\det(A - \lambda I) = 0$ is called the* ***characteristic equation*** *of A.*

Observation. *λ is an eigenvalue of a square matrix A if and only if λ satisfies the characteristic equation*

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

Example. Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

Solution.

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} 2 - \lambda & 3 \\ 3 & -6 - \lambda \end{bmatrix} \\ &= (2 - \lambda)(-6 - \lambda) - 9 \\ &= \lambda^2 + 4\lambda - 21 \\ &= (\lambda - 3)(\lambda + 7). \end{aligned}$$

Thus the eigenvalues of A are -7 and 3.

Example. Find the eigenvalues of $A = \begin{bmatrix} 3 & 3 & -1 \\ 0 & 2 & -6 \\ 0 & 0 & -1 \end{bmatrix}$.

Solution.

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

Thus the eigenvalues of A are 3, 2 and -1.

THEOREM 5.1. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

PROOF. Let A be a triangular matrix with diagonal entries $a_{11}, \ldots, a_{nn}$. Then $A-\lambda I$ is triangular matrix with diagonal entries $a_{11}-\lambda, \ldots, a_{nn}-\lambda$.

$$\det(A - I\lambda) = (a_{11} - \lambda)\ldots(a_{nn} - \lambda)$$

□

EXAMPLE. The eigenvalues of $\begin{bmatrix} -2 & 5 & 1 \\ 0 & 5 & 3 \\ 0 & 0 & 5 \end{bmatrix}$ are -2 and 5.

EXAMPLE. The eigenvalues of $\begin{bmatrix} 3 & 4 & -2 \\ 0 & 0 & 5 \\ 0 & 0 & 9 \end{bmatrix}$ are 3, 0 and 9.

What does it mean for a matrix A to have an eigenvalue of 0, as in this example?

$$\begin{aligned} 0 \text{ is an eigenvalue of } A &\iff Ax = 0 \text{ has a non-trivial solution} \\ &\iff \det A = 0 \\ &\iff A \text{ is not invertible.} \end{aligned}$$

This adds one more statement to the invertible matrix theorem.

THEOREM (The Invertible Matrix Theorem). *Let A be an $n \times n$ matrix. The following are equivalent.*

(a) A is an invertible matrix.
(b) $A \sim I_n$.
(c) A has n pivot positions.
(d) The equation $Ax = 0$ has only the trivial solution.
(e) The columns of A form a linearly independent set.
(f) The linear transformation $x \mapsto Ax$ is one-to-one.
(g) The equation $Ax = b$ has at least one solution for each $b \in \mathbb{R}^n$.
(h) The columns of A span $\mathbb{R}^n$.
(i) The linear transformation $x \mapsto Ax$ is onto.
(j) There is an $n \times n$ matrix C such that $CA = I_n$.
(k) There is an $n \times n$ matrix D such that $AD = I_n$.
(l) A^T is invertible.
(m) The columns of A form a basis for $\mathbb{R}^n$.
(n) $\operatorname{Col} A = \mathbb{R}^n$.
(o) $\operatorname{rank} A = n$.
(p) $\operatorname{nullity} A = 0$.
(q) $\operatorname{Nul} A = \{0\}$.
(r) The number 0 is not an eigenvalue of A.

Characteristic polynomial.

EXAMPLE. Find the characteristic equation of $A = \begin{bmatrix} 4 & 3 & -1 & 3 \\ 0 & 3 & 0 & 5 \\ 0 & 0 & -1 & 8 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.

SOLUTION.

$$\det(A - \lambda I) = (4-\lambda)(3-\lambda)(-1-\lambda)(3-\lambda) = (4-\lambda)(3-\lambda)^2(-1-\lambda).$$

So the characteristic equation of A is

$$(4-\lambda)(3-\lambda)^2(-1-\lambda) = 0$$

or

$$(\lambda - 4)(\lambda - 3)^2(\lambda + 1) = 0.$$

Expanding the product, we can also write

$$\lambda^4 - 9\lambda^3 + 23\lambda^2 - 3\lambda - 36 = 0.$$

Notice that $\det(A - \lambda I)$ is a polynomial in λ. It is called the *characteristic polynomial* of A. If A is $n \times n$, then its characteristic polynomial is of degree n.

In the example above, the characteristic polynomial is $(\lambda-4)(\lambda-3)^2(\lambda+1)$. We say that the eigenvalue 3 has multiplicty 2 because $(\lambda - 3)$ occurs two times as a factor. In general, the (*algebraic*) *multiplicity* of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

EXAMPLE. If the characteristic polynomial of a 6×6 matrix is $\lambda^6 - 2\lambda^5 - 15\lambda^4$, find the eigenvalues and their multiplicities.

SOLUTION.

$$\lambda^6 - 2\lambda^5 - 15\lambda^4 = \lambda^4(\lambda^2 - 2\lambda - 15) = \lambda^4(\lambda - 5)(\lambda + 3).$$

The eigenvalues are 0 (multiplicity 4), 5 (multiplicity 1) and -3 (multiplicity 1).

Similarity. Let A and B be $n \times n$ matrices. A is called *similar* to B if there is an invertible $n \times n$ matrix P such that

$$P^{-1}AP = B$$

or equivalently, $A = PBP^{-1}$.

If such P exists, then letting $Q = P^{-1}$, we have $Q^{-1}BQ = A$, so B is also similar to A. In short, we say that A and B are similar.

DEFINITION. *Let A and B be $n \times n$ matrices. A and B are* ***similar*** *if there is an invertible $n \times n$ matrix P such that*

$$P^{-1}AP = B.$$

A linear transformation of the form

$$\begin{aligned} M_{n\times n} &\to M_{n\times n} \\ A &\mapsto P^{-1}AP \end{aligned}$$

is called a similarity transformation.

THEOREM 5.4. *Let A and B be $n \times n$ matrices. If they are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).*

PROOF. If $P^{-1}AP = B$, then

$$\begin{aligned} B - \lambda I &= P^{-1}AP - \lambda I \\ &= P^{-1}AP - \lambda P^{-1}P \\ &= P^{-1}(A - \lambda I)P. \end{aligned}$$

Then

$$\det(B - \lambda I) = \det(P^{-1}) \det(A - \lambda I) \det P = \det(A - \lambda I).$$

□

REMARK.

- The matrices
$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$
are not similar even though they have the same eigenvalues.
- Similarity is not the same as row equivalence. The matrices
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$
are row equivalent but not similar (since their eigenvalues are not the same).

Cayley–Hamilton theorem. The characteristic polynomial of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$p(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = \lambda^2 + (-a - d)\lambda + (ad - bc).$$

Also notice that

$$\begin{aligned} A^2 + (-a-d)A + (ad-bc)I_2 &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}^2 + (-a-d) \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} \\ &= \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} + \begin{bmatrix} (-a-d)a & (-a-d)b \\ (-a-d)c & (-a-d)d \end{bmatrix} + \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} \\ &= \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} + \begin{bmatrix} -a^2-ad & -ab-bd \\ -ac-cd & -ad-d^2 \end{bmatrix} + \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

This property holds for any $n \times n$ matrix:

THEOREM (Cayley–Hamilton Theorem). *Let $p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$ be the characteristic polynomial of an $n \times n$ matrix A. Then*

$$A^n + c_{n-1}A^{n-1} + \ldots + c_1 A + c_0 I_n = 0$$

EXAMPLE. Recall that we found the characteristic polynomial of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$ as $\lambda^2 + 4\lambda - 21$. Hence

$$A^2 + 4A - 21I_2 = 0.$$

So,

$$A^2 = -4A + 21I_2.$$

Hence we can also calculate

$$\begin{aligned} A^3 &= A^2A = -4A^2 + 21A = -4(-4A + 21I_2) + 21A = 37A - 84I_2 \\ A^4 &= A^3A = 37A^2 - 84A = 37(-4A + 21I_2) - 84A = -232A + 777I_2 \end{aligned}$$

or

$$\begin{aligned} A^{-1} &= \frac{1}{21}(A + 4I_2) \\ A^{-2} &= A^{-1}A^{-1} = \frac{1}{441}(A^2 + 8A + 16I_2) = \frac{1}{441}(4A + 37I_2). \end{aligned}$$

5.3. Diagonalization

Powers of a diagonal matrices are easy to compute as in the following example.

EXAMPLE. If

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},$$

then

$$D^2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2^2 & 0 \\ 0 & 3^2 \end{bmatrix}$$

and

$$D^3 = DD^2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 2^3 & 0 \\ 0 & 3^3 \end{bmatrix}.$$

In general,

$$D^k = \begin{bmatrix} 2^k & 0 \\ 0 & 3^k \end{bmatrix}$$

for $k \geq 1$.

If $A = PDP^{-1}$ for some diagonal matrix D and invertible matrix P, then

$$A^2 = PDP^{-1}PDP^{-1} = PDDP^{-1} = PD^2P^{-1}$$

and

$$A^3 = AA^2 = PDP^{-1}PD^2P^{-1} = PDD^2P^{-1} = PD^3P^{-1}.$$

In general,

$$A^k = PD^kP^{-1}.$$

DEFINITION. *A square matrix A is said to be* **diagonalizable** *if A is similar to a diagonal matrix, that is, if*

$$A = PDP^{-1}$$

for some invertible matrix P and some diagonal matrix D.

EXAMPLE. Consider $P = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$ and $D = \begin{bmatrix} 7 & 0 \\ 0 & 4 \end{bmatrix}$.
Then, $P^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$ and $A = PDP^{-1} = \begin{bmatrix} 13 & 3 \\ -18 & -2 \end{bmatrix}$ is diagonalizable. We have

$$\begin{aligned} A^k &= PD^kP^{-1} \\ &= \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 7^k & 0 \\ 0 & 4^k \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \cdot 7^k - 2 \cdot 4^k & 7^k - 4^k \\ -6 \cdot 7^k + 6 \cdot 4^k & -2 \cdot 7^k + 3 \cdot 4^k \end{bmatrix}. \end{aligned}$$

THEOREM 5.5 (The Diagonalization Theorem). *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.*

In this case, if $v_1, v_2, \ldots, v_n$ are eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively, then $A = PDP^{-1}$, where

$$P = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an *eigenvector basis* of $\mathbb{R}^n$.

EXAMPLE. Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 5 & -8 & 12 \\ -8 & 5 & -12 \\ -8 & 8 & -15 \end{bmatrix}.$$

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

SOLUTION.

(1) **Find the eigenvalues of A:** We have

$$\begin{aligned}
\det(A-\lambda I) &= \det\begin{bmatrix} 5-\lambda & -8 & 12 \\ -8 & 5-\lambda & -12 \\ -8 & 8 & -15-\lambda \end{bmatrix} \\
&= \det\begin{bmatrix} 5-\lambda & -8 & 12 \\ -8 & 5-\lambda & -12 \\ -3-\lambda & 0 & -3-\lambda \end{bmatrix} \\
&= (-3-\lambda)\det\begin{bmatrix} -8 & 12 \\ 5-\lambda & -12 \end{bmatrix} + (-3-\lambda)\det\begin{bmatrix} 5-\lambda & -8 \\ -8 & 5-\lambda \end{bmatrix} \\
&= (-3-\lambda)(96-60+12\lambda) + (-3-\lambda)(\lambda^2-10\lambda+25-64) \\
&= (-3-\lambda)(12\lambda+36) + (-3-\lambda)(\lambda^2-10\lambda-39) \\
&= (-3-\lambda)(\lambda^2+2\lambda-3) \\
&= (-3-\lambda)(\lambda+3)(\lambda-1) \\
&= -(\lambda+3)^2(\lambda-1).
\end{aligned}$$

So the eigenvalues are $\lambda_1 = -3, \lambda_2 = -3$ and $\lambda_3 = 1$.

(2) **Find three linearly independent eigenvectors of A:** For $\lambda_1 = \lambda_2 = -3$, we need to solve

$$\begin{aligned}
\left[\begin{array}{c|c} A+3I & 0 \end{array}\right] &= \left[\begin{array}{ccc|c} 8 & -8 & 12 & 0 \\ -8 & 8 & -12 & 0 \\ -8 & 8 & -12 & 0 \end{array}\right] \\
&\sim \left[\begin{array}{ccc|c} 1 & -1 & 3/2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].
\end{aligned}$$

Its solution set is Span $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} \right\}$. Call

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}.$$

So $\{v_1, v_2\}$ is a basis for the eigenspace corresponding to $\lambda_1 = \lambda_2 = -3$.

For $\lambda_3 = 1$, we need to solve

$$\begin{aligned}
\left[\begin{array}{c|c} A-I & 0 \end{array}\right] &= \left[\begin{array}{ccc|c} 4 & -8 & 12 & 0 \\ -8 & 4 & -12 & 0 \\ -8 & 8 & -16 & 0 \end{array}\right] \\
&\sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].
\end{aligned}$$

Its solution set is Span $\left\{\begin{bmatrix}-1\\1\\1\end{bmatrix}\right\}$. Call

$$v_3 = \begin{bmatrix}-1\\1\\1\end{bmatrix}.$$

So $\{v_3\}$ is a basis for the eigenspace corresponding to $\lambda_3 = 1$.

One can check that $\{v_1, v_2, v_3\}$ is linearly independent. Thus A has 3 linearly independent eigenvectors. Hence, A is diagonalizable.

(3) **Construct P:**

$$P = \begin{bmatrix}v_1 & v_2 & v_3\end{bmatrix} = \begin{bmatrix}1 & -3 & -1\\1 & 0 & 1\\0 & 2 & 1\end{bmatrix}.$$

(4) **Construct D:**

$$D = \begin{bmatrix}\lambda_1 & 0 & 0\\0 & \lambda_2 & 0\\0 & 0 & \lambda_3\end{bmatrix} = \begin{bmatrix}-3 & 0 & 0\\0 & -3 & 0\\0 & 0 & 1\end{bmatrix}.$$

(5) **Check:** You can check that

$$AP = \begin{bmatrix}5 & -8 & 12\\-8 & 5 & -12\\-8 & 8 & -15\end{bmatrix}\begin{bmatrix}1 & -3 & -1\\1 & 0 & 1\\0 & 2 & 1\end{bmatrix} = \begin{bmatrix}-3 & 9 & -1\\-3 & 0 & 1\\0 & -6 & 1\end{bmatrix}$$

and

$$PD = \begin{bmatrix}1 & -3 & -1\\1 & 0 & 1\\0 & 2 & 1\end{bmatrix}\begin{bmatrix}-3 & 0 & 0\\0 & -3 & 0\\0 & 0 & 1\end{bmatrix} = \begin{bmatrix}-3 & 9 & -1\\-3 & 0 & 1\\0 & -6 & 1\end{bmatrix}.$$

So, $AP = PD$ and therefore $A = PDP^{-1}$.

Proof of theorem 5.5. ($\Rightarrow$): If A is diagonalizable, then $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. Then $AP = PD$. Call

$$P = \begin{bmatrix}v_1 & v_2 & \dots & v_n\end{bmatrix}$$

and

$$D = \begin{bmatrix}\lambda_1 & 0 & \dots & 0\\0 & \lambda_2 & \dots & 0\\\vdots & \vdots & & \vdots\\0 & 0 & \dots & \lambda_n\end{bmatrix}.$$

Then

$$AP = \begin{bmatrix}Av_1 & Av_2 & \dots & Av_n\end{bmatrix}$$

and

$$PD = \begin{bmatrix}\lambda_1 v_1 & \lambda_2 v_2 & \dots & \lambda_n v_n\end{bmatrix}.$$

Thus

$$Av_1 = \lambda_1 v_1, \quad Av_2 = \lambda_2 v_2, \quad \dots, \quad Av_n = \lambda_n v_n.$$

Hence $v_1, v_2, \ldots, v_n$ are eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. Since P is invertible, $\{v_1, v_2, \ldots, v_n\}$ is linearly independent.

$\underline{(\Leftarrow)}$: Let $v_1, v_2, \ldots, v_n$ be linearly independent eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. Then

$$P := \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

is invertible. Let

$$D := \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

Then

$$AP = \begin{bmatrix} Av_1 & Av_2 & \ldots & Av_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \ldots & \lambda_n v_n \end{bmatrix} = PD.$$

Since P is invertible, $A = PDP^{-1}$. □

Matrices Whose Eigenvalues Are Distinct.

THEOREM 5.2. *If $v_1, v_2, \ldots, v_r$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{v_1, v_2, \ldots, v_r\}$ is linearly independent.*

PROOF.

- First, we will show that:

$\{v_1, v_2, \ldots, v_r\}$ is linearly dependent $\implies$ $\{v_1, v_2, \ldots, v_{r-1}\}$ is linearly dependent.

Assume that $\{v_1, v_2, \ldots, v_r\}$ is linearly dependent. Then

$$x_1 v_1 + x_2 v_2 + \ldots + x_r v_r = 0$$

for some $x_1, x_2, \ldots, x_r \in \mathbb{R}$, not all zero, then

$$x_1 A v_1 + x_2 A v_2 + \ldots + x_r A v_r = A0 = 0$$

so

$$x_1 \lambda_1 v_1 + x_2 \lambda_2 v_2 + \ldots + x_r \lambda_r v_r = 0.$$

First and the last equations imply that

$$x_1(\lambda_1 - \lambda_r)v_1 + x_2(\lambda_2 - \lambda_r)v_2 + \ldots + x_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1} = 0.$$

If $\{v_1, v_2, \ldots, v_{r-1}\}$ is linearly independent, then for each $1 \le i \le r-1$, we have $x_i(\lambda_i - \lambda_r) = 0$. Since eigenvalues are distinct, $\lambda_i - \lambda_r \neq 0$, so

$$x_1 = x_2 = \ldots = x_{r-1} = 0.$$

Then $x_r \neq 0$ and $x_r v_r = 0$. Then $v_r = 0$. But $v_r \neq 0$ since v_r is an eigenvector. Thus we get a contradiction So $\{v_1, v_2, \ldots, v_{r-1}\}$ is linearly dependent.

- Continuing the same argument again and again:

$$
\begin{aligned}
\{v_1, v_2, \ldots, v_r\} \text{ is linearly dependent } &\Longrightarrow \{v_1, v_2, \ldots, v_{r-1}\} \text{ is linearly dependent} \\
&\Longrightarrow \{v_1, v_2, \ldots, v_{r-2}\} \text{ is linearly dependent} \\
&\Longrightarrow \{v_1, v_2, \ldots, v_{r-3}\} \text{ is linearly dependent} \\
&\vdots \\
&\Longrightarrow \{v_1, v_2\} \text{ is linearly dependent} \\
&\Longrightarrow \{v_1\} \text{ is linearly dependent} \\
&\Longrightarrow v_1 = 0.
\end{aligned}
$$

  This is a contradiction, since v_1 is an eigenvector. Thus $\{v_1, v_2, \ldots, v_r\}$ is linearly independent.

$\square$

By theorem 5.5, this implies the following.

THEOREM 5.6. *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

EXAMPLE. Determine if the following matrix is diagonalizable.

$$
A = \begin{bmatrix} 2 & -1 & -3 \\ 0 & 0 & 1 \\ 0 & 0 & -7 \end{bmatrix}.
$$

SOLUTION. Eigenvalues of A are 2, 0 and -7. Since A is a 3×3 matrix with 3 distinct eigenvalues, A is diagonalizable.

Matrices Whose Eigenvalues Are Not Distinct.

EXAMPLE. Diagonalize the following matrix, if possible.

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

SOLUTION. $\det(A - \lambda I) = -(\lambda - 1)(\lambda + 2)^2$. So the eigenvalues of A are 1, -2 and -2.

One can calculate that

$$
\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}
$$

is a basis for the eigenspace corresponding to $\lambda_1 = 1$ and

$$
\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}
$$

is a basis for the eigenspace corresponding to $\lambda_2 = \lambda_3 = -2$.

Call $v_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$. Any other eigenvector of A is either multiple of v_1 or v_2. Hence it is impossible to construct a basis of $\mathbb{R}^3$ using eigenvectors of A. Thus A is not diagonalizable by theorem 5.5.

THEOREM 5.7. *Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_r$.*

(i) For $1 \leq i \leq r$,

$$\dim(\textit{eigenspace for } \lambda_i) \leq \textit{multiplicity of the eigenvalue } \lambda_i.$$

(ii) The following are equivalent.
  - *(a) A is diagonalizable*
  - *(b) the sum of the dimensions of the eigenspaces equals n*
  - *(c) the characteristic polynomial factors completely into linear factors and for $1 \leq i \leq r$,*

$$\dim(\textit{eigenspace for } \lambda_i) = \textit{multiplicity of the eigenvalue } \lambda_i.$$

(iii) If A is diagonalizable and $\mathcal{B}_i$ is a basis for the eigenspace corresponding to λ_i for each i, then the total collection of vectors in the sets $\mathcal{B}_1, \ldots, \mathcal{B}_r$ forms an eigenvector basis for $\mathbb{R}^n$.

We omit the proof in our course. Curious reader may read the proof in the book Linear algebra by Friedberg, Insel, Spence (4th edition), section 5.2.

EXAMPLE. Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 1 & 4 & -3 & 0 \\ -1 & -2 & 0 & -3 \end{bmatrix}.$$

SOLUTION. $\det(A - \lambda I) = (\lambda - 5)^2(\lambda + 3)^2$. So the eigenvalues of A are $5, 5, -3, -3$. One can calculate that

$$v_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

form a basis for the eigenspace corresponding to $\lambda_1 = \lambda_2 = 5$ and

$$v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, v_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

form a basis for the eigenspace corresponding to $\lambda_3 = \lambda_4 = -3$.

Then by theorem 5.7, A is diagonalizable.

Therefore, putting

$$P = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix} = \begin{bmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

we have

$$A = PDP^{-1}.$$

CHAPTER 6

Orthogonality

6.1. Inner Product, Length, and Orthogonality

The Inner Product.

DEFINITION. *Let*

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n.$$

The ***standard inner product*** *(or* ***dot product****) of* u *and* v *is defined as*

$$u^T v = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \ldots + u_n v_n$$

and denoted as $u \cdot v$.

EXAMPLE.

$$\begin{bmatrix} 6 \\ 2 \\ -7 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -4 \\ -1 \end{bmatrix} = 6 \cdot 3 + 2 \cdot (-4) + (-7) \cdot (-1) = 18 - 8 + 7 = 17.$$

Following results follow directly from the definition.

THEOREM 6.1. *Let* $u, v, w \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$. *Then*

(a) $u \cdot v = v \cdot u$
(b) $(u + w) \cdot v = u \cdot v + w \cdot v$
(c) $(cu) \cdot v = c(u \cdot v)$
(d) $u \cdot u \geq 0$
(e) $u \cdot u = 0 \iff u = 0$.

(b) and (c) implies the following.

COROLLARY. *Let* $v, u_i \in \mathbb{R}^n$ *and* $c_i \in \mathbb{R}$. *Then*

$$(c_1 u_1 + \ldots + c_m u_m) \cdot v = c_1(u_1 \cdot v) + \ldots c_m(u_m \cdot v).$$

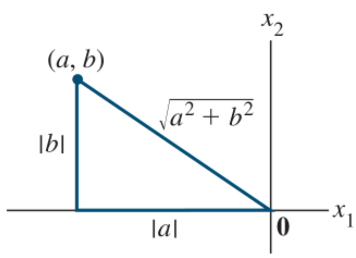
The Length of a Vector.

DEFINITION. *The* **length** *(or* **norm***) of* $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ *is the nonnegative scalar* $\|v\|$ *defined by*

$$\|v\| := \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

EXAMPLE. If $v = \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$,

$$\|v\| = \sqrt{a^2 + b^2}.$$



OBSERVATION. *For any* $c \in \mathbb{R}$ *and* $v \in \mathbb{R}^n$,

$$\|cv\| = |c|\|v\|.$$

PROOF.

$$\|cv\|^2 = (cv) \cdot (cv) = c^2(v \cdot v) = c^2\|v\|^2$$
$$\implies \|cv\| = |c|\|v\|.$$

□

DEFINITION. *A vector whose length is 1 is called a* **unit vector***.*

If v is a non-zero vector, the length of

$$u = \frac{1}{\|v\|} v$$

is

$$\|u\| = \frac{1}{\|v\|}\|v\| = 1.$$

The process of creating u from v is called *normalizing* v. We say that u is the *unit vector in the same direction as* v.

EXAMPLE. Find a unit vector u in the same direction as $v = \begin{bmatrix} -4 \\ 2 \\ 1 \\ -2 \\ 0 \end{bmatrix}$.

SOLUTION.
$$\|v\| = \sqrt{(-4)^2 + 2^2 + 1^2 + (-2)^2 + 0^2} = \sqrt{16+4+1+4+0} = \sqrt{25} = 5.$$
Then
$$u = \frac{1}{\|v\|}v = \frac{1}{5}\begin{bmatrix} -4 \\ 2 \\ 1 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -4/5 \\ 2/5 \\ 1/5 \\ -2/5 \\ 0 \end{bmatrix}.$$

EXAMPLE. Let $W = \text{Span}\left\{\begin{bmatrix} 5/12 \\ 1 \end{bmatrix}\right\}$. Find a unit vector u such that $\{u\}$ is a basis for W.

SOLUTION. We have $W = \text{Span}\left\{\begin{bmatrix} 5/12 \\ 1 \end{bmatrix}\right\} = \text{Span}\left\{\begin{bmatrix} 5 \\ 12 \end{bmatrix}\right\}$. Call $v = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$. Normalizing v, we have
$$u = \frac{1}{\|v\|}v = \frac{1}{13}\begin{bmatrix} 5 \\ 12 \end{bmatrix} = \begin{bmatrix} 5/13 \\ 12/13 \end{bmatrix}.$$
$W = \text{Span}\{v\} = \text{Span}\{u\}$, so $\{u\}$ is a basis for W.
Another unit vector giving a basis for W is $\begin{bmatrix} -5/13 \\ -12/13 \end{bmatrix}$.

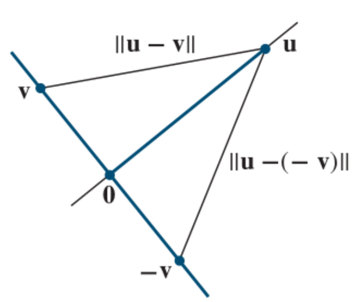
Distance.

DEFINITION. *Let $u, v \in \mathbb{R}^n$. The **distance** between u and v is defined as*
$$\text{dist}(u, v) = \|u - v\|.$$

EXAMPLE. Find the distance between $u = (2, 3)$ and $v = (-1, 5)$.

SOLUTION.
$$u - v = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$
So,
$$\text{dist}(u, v) = \|u - v\| = \sqrt{3^2 + (-2)^2} = \sqrt{13}.$$

Orthogonal Vectors. Consider the following vectors u, v.



Recall from highschool that

$$u \text{ and } v \text{ are perpendicular} \iff \mathsf{dist}(u, v) = \mathsf{dist}(u, -v).$$

We have

$$\begin{aligned}\mathsf{dist}(u, v)^2 &= \|u - v\|^2 \\ &= (u - v) \cdot (u - v) \\ &= u \cdot u - u \cdot v - v \cdot u + v \cdot v \\ &= \|u\|^2 + \|v\|^2 - 2u \cdot v.\end{aligned}$$

and

$$\begin{aligned}\mathsf{dist}(u, -v)^2 &= \|u - (-v)\|^2 \\ &= \|u + v\|^2 \\ &= (u + v) \cdot (u + v) \\ &= u \cdot u + u \cdot v + v \cdot u + v \cdot v \\ &= \|u\|^2 + \|v\|^2 + 2u \cdot v.\end{aligned}$$

Thus

$$\begin{aligned}u \text{ and } v \text{ are perpendicular} &\iff \mathsf{dist}(u, v) = \mathsf{dist}(u, -v) \\ &\iff \|u\|^2 + \|v\|^2 - 2u \cdot v = \|u\|^2 + \|v\|^2 + 2u \cdot v \\ &\iff 4u \cdot v = 0 \\ &\iff u \cdot v = 0.\end{aligned}$$

We call 'orthogonal' instead of 'perpendicular' in linear algebra.

DEFINITION. *Two vectors u and v in $\mathbb{R}^n$ are called* ***orthogonal*** *(to each other) if*

$$u \cdot v = 0.$$

By the calculations above,

$$\begin{aligned}u \text{ and } v \text{ are orthogonal} &\iff u \cdot v = 0 \\ &\iff \|u + v\|^2 = \|u\|^2 + \|v\|^2.\end{aligned}$$

This is the famous The Pythagorean Theorem.

THEOREM 6.2 (The Pythagorean Theorem). *Two vectors u and v are orthogonal if and only if $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

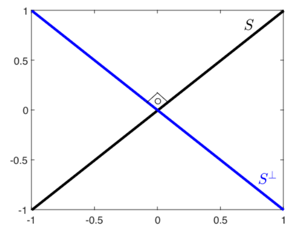
Orthogonal Complements.

DEFINITION. *If a vector v is orthogonal to every vector in a subspace S of $\mathbb{R}^n$, then v is said to be* ***orthogonal*** *to S.*

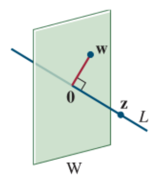
DEFINITION. *Let S be a subspace of $\mathbb{R}^n$. The* ***orthogonal complement*** *of S, denoted by $S^\perp$, is the subset of $\mathbb{R}^n$ consisting of all vectors that are orthogonal to S, that is*

$$S^\perp = \{v \in \mathbb{R}^n \mid v \cdot u = 0, \quad \forall u \in S\}.$$

EXAMPLE. In $\mathbb{R}^2$, two orthogonal lines through 0 are orthogonal complements of each other.



EXAMPLE. In $\mathbb{R}^3$, a plane W through 0 and line L through 0 that is perpendicular to W are orthogonal complements of each other.



EXAMPLE. Let

$$S = \text{Span}\left\{\begin{bmatrix}1\\2\\-1\end{bmatrix}\right\}.$$

Then,

$$\begin{aligned}
S^{\perp} &= \left\{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \mid \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \cdot \left(c\begin{bmatrix}1\\2\\-1\end{bmatrix}\right) = 0, \quad \forall c \in \mathbb{R}\right\} \\
&= \left\{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \mid c(x_1 + 2x_2 - x_3) = 0, \quad \forall c \in \mathbb{R}\right\} \\
&= \left\{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \mid x_1 + 2x_2 - x_3 = 0\right\} \\
&= \left\{\begin{bmatrix}x_1\\x_2\\x_1 + 2x_2\end{bmatrix} \mid x_1, x_2 \in \mathbb{R}\right\} \\
&= \text{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}\right\}.
\end{aligned}$$

OBSERVATION. *Let $S = \text{Span}\{v_1, \ldots, v_m\}$.*

$$u \in S^{\perp} \iff u \cdot v_i = 0, \quad \forall i \in \{1, \ldots, m\}.$$

OBSERVATION. *Let S be a subspace of $\mathbb{R}^n$. Then $S^{\perp}$ is a subspace of $\mathbb{R}^n$.*

Proof of this observation is left as an exercise.

THEOREM 6.3. *Let A be an $m \times n$ matrix. Then*

$$(\operatorname{Row} A)^{\perp} = \operatorname{Nul} A.$$

EXAMPLE. Let

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

Then,

$$\operatorname{Row} A = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

and

$$\operatorname{Nul} A = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}.$$

PROOF. Call r_i for the r-th row of A. Then $\operatorname{Row} A = \operatorname{Span}\{r_1, \ldots, r_m\}$. Hence

$$\begin{aligned} v \in \operatorname{Nul} A &\iff Av = 0 \\ &\iff r_i \cdot v = 0, \quad \forall i \in \{1, \ldots, m\} \\ &\iff u \cdot v = 0, \quad \forall u \in \operatorname{Row} A \\ &\iff v \in (\operatorname{Row} A)^{\perp}. \end{aligned}$$

□

COROLLARY.

$$(\operatorname{Col} A)^{\perp} = \operatorname{Nul} A^{T}.$$

6.2. Orthogonal Sets

DEFINITION. *A set of vectors $\{v_1, \ldots, v_m\}$ in $\mathbb{R}^n$ consisting of nonzero vectors is called an* **orthogonal set** *if each pair of distinct vectors from the set is orthogonal, that is, if*

$$v_i \cdot v_j = 0$$

whenever $i \neq j$.

EXAMPLE. Let

$$v_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, v_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}.$$

Then

$$\begin{aligned} v_1 \cdot v_2 &= 3 \cdot 1 + 1 \cdot (-2) + 1 \cdot (-1) = 3 - 2 - 1 = 0 \\ v_1 \cdot v_3 &= 3 \cdot (-1) + 1 \cdot (-4) + 1 \cdot 7 = -3 - 4 + 7 = 0 \\ v_2 \cdot v_3 &= 1 \cdot (-1) + (-2) \cdot (-4) + (-1) \cdot 7 = -1 + 8 - 7 = 0. \end{aligned}$$

Hence $\{v_1, v_2, v_3\}$ is an orthogonal set.

THEOREM 6.4. *Any orthongal set is linearly independent.*

PROOF. Let $\{v_1, \ldots, v_m\}$ be an orthogonal set. If

$$c_1 v_1 + \ldots + c_m v_m = 0,$$

then for any i,

$$\begin{aligned} 0 = 0 \cdot v_i &= (c_1 v_1 + \ldots + c_m v_m) \cdot v_i \\ &= c_1(v_1 \cdot v_i) + \ldots + c_m(v_m \cdot v_i) \\ &= c_i(v_i \cdot v_i). \end{aligned}$$

Since $v_i \neq 0$, we have $v_i \cdot v_i \neq 0$. Then $c_i = 0$. □

COROLLARY.

(i) An orthongal set $S = \{v_1, \ldots, v_m\}$ in $\mathbb{R}^n$ is a basis for Span S.

(ii) An orthongal set $S = \{v_1, \ldots, v_n\}$ in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$.

REMARK. A linearly independent set does not have to be orthogonal.

EXAMPLE. $\left\{ \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix} \right\}$ is orthogonal, so it is also linearly independent.

EXAMPLE. $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is linearly independent, but not orthogonal.

DEFINITION. *An* ***orthogonal basis*** *for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.*

EXAMPLE. The standard basis $\{e_1, \ldots, e_n\}$ is an orthogonal basis for $\mathbb{R}^n$.

EXAMPLE. $\left\{ \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^3$.

THEOREM 6.5. *Let $\{v_1, \ldots, v_m\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For every $w \in W$, if*

$$w = c_1 v_1 + \ldots + c_m v_m,$$

then

$$c_i = \frac{w \cdot v_i}{v_i \cdot v_i}$$

for all $i = 1, \ldots, m$.

The theorem says the following. If $\mathcal{B} = \{v_1, \ldots, v_m\}$ is an orthogonal basis for W, then for every $w \in W$,

$$w = \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + \ldots + \frac{w \cdot v_m}{v_m \cdot v_m} v_m.$$

So

$$[w]_{\mathcal{B}} = \begin{bmatrix} \frac{w \cdot v_1}{v_1 \cdot v_1} \\ \frac{w \cdot v_2}{v_2 \cdot v_2} \\ \vdots \\ \frac{w \cdot v_m}{v_m \cdot v_m} \end{bmatrix}.$$

PROOF. If

$$w = c_1 v_1 + \ldots + c_m v_m,$$

then by applying inner product with v_i

$$w \cdot v_i = c_i(v_i \cdot v_i).$$

□

EXAMPLE. Let

$$v_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, v_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}$$

as in the example above. Then $S = \{v_1, v_2, v_3\}$ is an orthogonal basis for $\mathbb{R}^3$. Let us express the vector $w = \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix}$ as a linear combination of the vectors in S.

$$\begin{aligned} w \cdot v_1 &= 19, \quad w \cdot v_2 = 3, \quad w \cdot v_3 = -21, \\ v_1 \cdot v_1 &= 11, \quad v_2 \cdot v_2 = 6, \quad v_3 \cdot v_3 = 66. \end{aligned}$$

Thus

$$\begin{aligned} w &= \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + \frac{w \cdot v_3}{v_3 \cdot v_3} v_3 \\ &= \frac{19}{11} v_1 + \frac{3}{6} v_2 + \frac{-21}{66} v_3 \\ &= \frac{19}{11} v_1 + \frac{1}{2} v_2 + \frac{-7}{22} v_3. \end{aligned}$$

Orthonormal Sets.

DEFINITION.

- *An orthogonal set of unit vectors is called an **orthonormal set**.*
- *An **orthonormal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthonormal set.*

EXAMPLE. The standard basis $\{e_1, \ldots, e_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

EXAMPLE. $\left\{ \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^3$.

THEOREM 6.6. *An $m \times n$ matrix U has orthonormal columns if and only if*

$$U^T U = I_n.$$

PROOF. Call $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$. We have

$$\begin{aligned} U^T U &= \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \\ &= \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \dots & u_1^T u_n \\ u_2^T u_1 & u_2^T u_2 & \dots & u_2^T u_n \\ \vdots & \vdots & \vdots & \vdots \\ u_n^T u_1 & u_n^T u_2 & \dots & u_n^T u_n \end{bmatrix} \\ &= \begin{bmatrix} u_1 \cdot u_1 & u_1 \cdot u_2 & \dots & u_1 \cdot u_n \\ u_2 \cdot u_1 & u_2 \cdot u_2 & \dots & u_2 \cdot u_n \\ \vdots & \vdots & \vdots & \vdots \\ u_n \cdot u_1 & u_n \cdot u_2 & \dots & u_n \cdot u_n \end{bmatrix}. \end{aligned}$$

Thus

$$\begin{aligned} U \text{ has orthonormal columns } &\iff \{u_1, \dots, u_n\} \text{ is an orthonormal set} \\ &\iff u_i \cdot u_j = 0, \text{ when } i \neq j, \text{ and } u_i \cdot u_i = 1 \\ &\iff U^T U = I_n. \end{aligned}$$

□

EXAMPLE. Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$. Then

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so U has orthonormal columns.

THEOREM 6.7. *Let U be an $m \times n$ matrix with orthonormal columns, and $x, y \in \mathbb{R}^n$. Then $(Ux) \cdot (Uy) = x \cdot y$.*

PROOF. $(Ux) \cdot (Uy) = (Ux)^T (Uy) = x^T U^T U y = x^T y = x \cdot y$. □

Taking $x = y$ in the theorem above:

COROLLARY. $\|Ux\| = \|x\|$.

6.4. The Gram Schmidt Process

The Gram–Schmidt process is a simple algorithm for producing an orthogonal basis for any nontrivial subspace of $\mathbb{R}^n$.

Let $\{v_1, v_2, \dots, v_m\}$ be a basis for a nontrivial subspace W of $\mathbb{R}^n$. We would like to find an orthogonal basis

$$\{u_1, u_2, \dots u_m\}$$

for

$$W = \text{Span}\{v_1, v_2, \ldots, v_m\}.$$

- First, let us find an orthogonal basis

$$\{u_1, u_2\}$$

for

$$W_2 = \text{Span}\{v_1, v_2\}.$$

Set

$$u_1 = v_1.$$

Choose u_2 based on

$$v_2 = \alpha_1 u_1 + u_2.$$

Now

$$\begin{aligned} u_1 \text{ and } u_2 \text{ are orthongal} &\iff u_2 \cdot u_1 = 0 \\ &\iff v_2 \cdot u_1 = \alpha_1 u_1 \cdot u_1 \\ &\iff \alpha_1 = \frac{v_2 \cdot u_1}{u_1 \cdot u_1}. \end{aligned}$$

Hence we shall choose

$$\alpha_1 = \frac{v_2 \cdot u_1}{u_1 \cdot u_1}$$

and

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1.$$

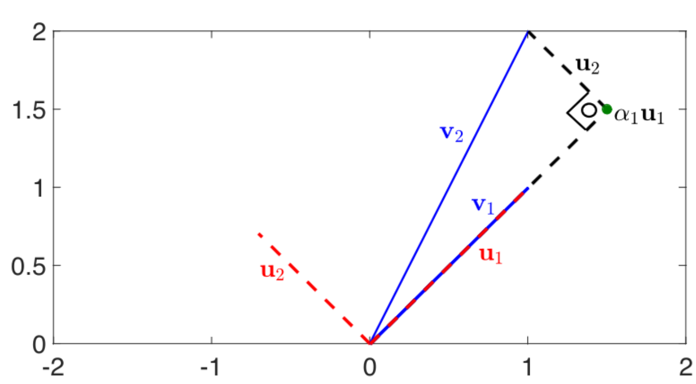
Thus:

Observation. *Taking*

$$u_1 = v_1, \quad u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1$$

the set $\{u_1, u_2\}$ *is an orthogonal basis for* $W_2 = \text{Span}\{v_1, v_2\}$.

Illustration of how we choose u_2:



- Now, suppose we have an orthogonal basis

$$\{u_1, \ldots, u_{q-1}\}$$

for

$$W_{q-1} = \text{Span}\{v_1, \ldots, v_{q-1}\}.$$

Let us find an orthogonal basis

$$\{u_1, \ldots, u_{q-1}, u_q\}$$

for

$$W_q = \text{Span}\{v_1, \ldots, v_{q-1}, v_q\}.$$

Choose u_q based on

$$v_q = \alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_{q-1} u_{q-1} + u_q.$$

Now for each $j \in \{1, \ldots, q-1\}$,

$$\begin{aligned} u_q \text{ is orthogonal to } u_j &\iff u_q \cdot u_j = 0 \\ &\iff v_q \cdot u_j = \alpha_j u_j \cdot u_j \\ &\iff \alpha_j = \frac{v_q \cdot u_j}{u_j \cdot u_j}. \end{aligned}$$

Hence we shall choose

$$\alpha_j = \frac{v_q \cdot u_j}{u_j \cdot u_j} \quad \text{for } j = 1, \ldots, q-1$$

and

$$u_q = v_q - \frac{v_q \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_q \cdot u_2}{u_2 \cdot u_2} u_2 - \ldots - \frac{v_q \cdot u_{q-1}}{u_{q-1} \cdot u_{q-1}} u_{q-1}.$$

Thus

$$\{u_1, \ldots, u_q\}$$

is an orthogonal basis for

$$W_q = \text{Span}\{v_1, \ldots, v_q\}.$$

We showed the following.

Theorem 6.11 (The Gram–Schmidt Process). *Let W be a subspace of $\mathbb{R}^n$ with a basis $\{v_1, \ldots, v_m\}$. Then the set $\{u_1, \ldots, u_m\}$ defined by*

$$u_1 = v_1$$

and

$$u_q = v_q - \frac{v_q \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_q \cdot u_2}{u_2 \cdot u_2} u_2 - \ldots - \frac{v_q \cdot u_{q-1}}{u_{q-1} \cdot u_{q-1}} u_{q-1}$$

for $q = 2, 3, \ldots, m$, is an orthogonal basis for W.

Example. Find an orthogonal basis for the subspace W of $\mathbb{R}^4$ with the basis $\{v_1, v_2, v_3\}$, where

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 5 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 5 \end{bmatrix}.$$

Solution. Choose as in the Gram–Schmidt process

$$u_1 = v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 5 \end{bmatrix} - \frac{8}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 3 \end{bmatrix}$$

$$u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 5 \end{bmatrix} - \frac{4}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{15}{14} \begin{bmatrix} -2 \\ -1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 8/7 \\ 1/14 \\ -2 \\ 11/14 \end{bmatrix}.$$

Hence

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 8/7 \\ 1/14 \\ -2 \\ 11/14 \end{bmatrix} \right\}$$

is an orthogonal basis for W.

6.3. Orthogonal Projections

Orthogonal Decomposition. Let W be an m-dimensional subspace of $\mathbb{R}^n$. Let

$$\mathcal{U} = \{u_1, \ldots, u_m\}$$

be an orthogonal basis for W (it can be formed using the Gram-Schmidt process).

By applying the Gram-Schmidt process, extend the basis $\mathcal{U}$ to an orthogonal basis

$$\{u_1, \ldots, u_m, u_{m+1}, \ldots, u_n\}$$

for $\mathbb{R}^n$.

Exercise. Show that

$$W^\perp = \operatorname{Span}\{u_{m+1}, \ldots, u_n\}.$$

Thus, any vector $v \in \mathbb{R}^n$ can be written as

$$v = \underbrace{\alpha_1 u_1 + \ldots + \alpha_m u_m}_{\widehat{v}} + \underbrace{\alpha_{m+1} u_{m+1} + \ldots + \alpha_n u_n}_{p}$$

where $\widehat{v} \in W$ and $p \in W^\perp$.

Moreover, such a decomposition of v is unique: If $v = \widehat{v}_0 + p_0$, for some other $\widehat{v}_0 \in W$ and $p_0 \in W^{\perp}$, then

$$\begin{aligned} v = \widehat{v} + p = \widehat{v}_0 + p_0 &\implies \widehat{v} - \widehat{v}_0 = p_0 - p \in W^{\perp} \\ &\implies \widehat{v} - \widehat{v}_0 \in W, \quad \widehat{v} - \widehat{v}_0 \in W^{\perp} \\ &\implies (\widehat{v} - \widehat{v}_0) \cdot (\widehat{v} - \widehat{v}_0) = 0 \\ &\implies \|\widehat{v} - \widehat{v}_0\| = 0 \\ &\implies \widehat{v} - \widehat{v}_0 = 0 \\ &\implies \widehat{v} = \widehat{v}_0 \text{ and } p = p_0. \end{aligned}$$

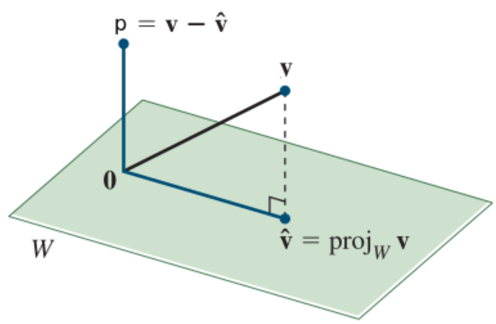
We showed the following.

THEOREM 6.8 (The Orthogonal Decomposition Theorem). *Let W be a subspace of $\mathbb{R}^n$. Every $v \in \mathbb{R}^n$ can be written uniquely in the form*

$$v = \widehat{v} + p$$

where $\widehat{v} \in W$ and $p \in W^{\perp}$.

The vector $\operatorname{proj}_W v := \widehat{v}$ in the theorem is called the *orthogonal projection* of y onto W.



Note that if $v \in W$, then $\operatorname{proj}_W v = v$.

Computation of Orthogonal Projection.

EXAMPLE. Let

$$u_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \quad v = \begin{bmatrix} 10 \\ 0 \\ 9 \end{bmatrix}, \quad W = \operatorname{Span}\{u_1\}.$$

Find the orthogonal projection of v onto W.

SOLUTION. Let u_2, u_3 be vectors such that $\{u_1, u_2, u_3\}$ is an orthogonal basis for $\mathbb{R}^3$. Then

$$v = \underbrace{\alpha_1 u_1}_{\widehat{v}} + \underbrace{\alpha_2 u_2 + \alpha_3 u_3}_{p}$$

where

$$\alpha_1 = \frac{v \cdot u_1}{u_1 \cdot u_1}, \quad \alpha_2 = \frac{v \cdot u_2}{u_2 \cdot u_2}, \quad \alpha_3 = \frac{v \cdot u_3}{u_3 \cdot u_3}$$

by theorem 6.5. Thus

$$\operatorname{proj}_W v = \widehat{v} = \frac{v \cdot u_1}{u_1 \cdot u_1} u_1 = \frac{28}{14} \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -6 \\ 4 \end{bmatrix}.$$

Also

$$v - \widehat{v} = \begin{bmatrix} 8 \\ 6 \\ 5 \end{bmatrix} \in W^{\perp}.$$

We may generalize this idea.

THEOREM 6.10. *Let W be a subspace of $\mathbb{R}^n$ and $v \in \mathbb{R}^n$.*

- *If $\{u_1, \ldots, u_m\}$ is an orthogonal basis for W,*

$$\operatorname{proj}_W v = \frac{v \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{v \cdot u_2}{u_2 \cdot u_2} u_2 + \ldots + \frac{v \cdot u_m}{u_m \cdot u_m} u_m.$$

- *If $\{u_1, \ldots, u_m\}$ is an orthonormal basis for W,*

$$\operatorname{proj}_W v = (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + \ldots + (v \cdot u_m)u_m.$$

Also, letting $U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix}$,

$$\operatorname{proj}_W v = UU^T v.$$

PROOF.

- Let W be a subspace of $\mathbb{R}^n$ and

$$\mathcal{U} = \{u_1, \ldots, u_m\}$$

be an orthogonal basis for W. Extend $\mathcal{U}$ to an orthogonal basis

$$\{u_1, \ldots, u_m, u_{m+1}, \ldots, u_n\}$$

for $\mathbb{R}^n$. We have

$$v = \underbrace{\alpha_1 u_1 + \ldots + \alpha_m u_m}_{\widehat{v}} + \underbrace{\alpha_{m+1} u_{m+1} + \ldots + \alpha_n u_n}_{p}$$

with

$$\alpha_j = \frac{v \cdot u_j}{u_j \cdot u_j}$$

for $j = 1, \ldots, n$, by theorem 6.5.

- If $\mathcal{U}$ is moreover orthonormal, then $u_1 \cdot u_1 = \ldots = u_m \cdot u_m = 1$, therefore

$$\begin{aligned}
\operatorname{proj}_W v &= (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + \ldots + (v \cdot u_m)u_m \\
&= \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \begin{bmatrix} u_1 \cdot v \\ u_2 \cdot v \\ \vdots \\ u_m \cdot v \end{bmatrix} \\
&= \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} v \\
&= UU^T v.
\end{aligned}$$

$\square$

Example. Let

$$u_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix}, \quad v = \begin{bmatrix} -2 \\ 5 \\ -11 \end{bmatrix}, \quad W = \operatorname{Span}\{u_1, u_2\}.$$

Then $\{u_1, u_2\}$ is an orthogonal basis for W and

$$\begin{aligned}
\operatorname{proj}_W v &= \frac{v \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{v \cdot u_2}{u_2 \cdot u_2} u_2 \\
&= \frac{1}{5} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{-32}{21} \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} 202/105 \\ -299/105 \\ -128/21 \end{bmatrix}.
\end{aligned}$$

Let U be a matrix whose columns is an orthonormal basis for a subspace W of $\mathbb{R}^n$.

- $U^T U = I$.
- The *orthogonal projection transformation* $T(v) = \operatorname{proj}_W v$ is linear with the standard matrix UU^T.
- UU^T is called the *orthogonal projector* onto W.

Example. Let $W = \operatorname{Span}\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix} \right\}$.

The orthogonal projector onto W is

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 17/18 & 1/18 & 2/9 \\ 1/18 & 17/18 & -2/9 \\ 2/9 & -2/9 & 1/9 \end{bmatrix}.$$

For every $v \in \mathbb{R}^3$,

$$\operatorname{proj}_W v = \begin{bmatrix} 17/18 & 1/18 & 2/9 \\ 1/18 & 17/18 & -2/9 \\ 2/9 & -2/9 & 1/9 \end{bmatrix} v.$$

The Best Approximation Problem. The best approximation problem is given

- a subspace W of $\mathbb{R}^n$
- a point (vector) $v \in \mathbb{R}^n$

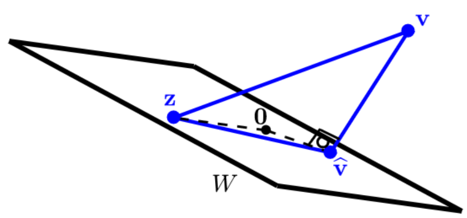
finding

- $x \in W$ such that $\operatorname{dist}(v, x) = \|v - x\|$ is minimal
  (i.e. $\operatorname{dist}(v, x) < \operatorname{dist}(v, z)$ for any $z \in W$ such that $z \neq x$).

Let us consider the orthogonal projection $\widehat{v} = \operatorname{proj}_W v$. For any $z \in W$, with $z \neq \widehat{v}$, we have

$$v - z = \underbrace{(v - \widehat{v})}_{\in W^\perp} + \underbrace{(\widehat{v} - z)}_{\in W}.$$

So $\widehat{v} - z = \operatorname{proj}_W(v - z)$.



By the Pythagorean Theorem,

$$\|v - \widehat{v}\|^2 + \|\widehat{v} - z\|^2 = \|v - z\|^2.$$

Thus

$$\|v - \widehat{v}\| < \|v - z\|.$$

Theorem 6.9 (The Best Approximation Theorem). *Let W be a subspace of $\mathbb{R}^n$ and $v \in \mathbb{R}^n$. Then we have*

$$\|v - \operatorname{proj}_W v\| < \|v - z\|$$

for all $z \in W$ such that $z \neq \operatorname{proj}_W v$.

$\|v - \operatorname{proj}_W v\|$ is called the *distance of v to the subspace W.*