LINEAR ALGEBRA MATH 2700.006 SPRING 2013 (COHEN) LECTURE NOTES

1 Sets and Set Notation.

Definition 1 (Naive Definition of a Set)**.** A **set** is any collection of objects, called the **elements** of that set. We will most often name sets using capital letters, like A, B, X, Y, etc., while the elements of a set will usually be given lower-case letters, like x, y, z, v, etc.

Two sets X and Y are called **equal** if X and Y consist of exactly the same elements. In this case we write $X = Y$.

Example 1 (Examples of Sets)**.**

(1) Let X be the collection of all *integers* greater than or equal to 5 and strictly less than 10. Then X is a set, and we may write:

$$X = \{5, 6, 7, 8, 9\}$$

The above notation is an example of a set being described **explicitly**, i.e. just by listing out all of its elements. The **set brackets** $\{\cdots\}$ indicate that we are talking about a set and not a number, sequence, or other mathematical object.

(2) Let E be the set of all even *natural numbers*. We may write:

$$E = \{0, 2, 4, 6, 8, ...\}$$

This is an example of an explicity described set with infinitely many elements. The ellipsis (...) in the above notation is used somewhat informally, but in this case its meaning, that we should "continue counting forever," is clear from the context.

(3) Let Y be the collection of all *real numbers* greater than or equal to 5 and strictly less than 10. Recalling notation from previous math courses, we may write:

$$Y = [5, 10)$$

This is an example of using **interval notation** to describe a set. Note that the set Y obviously consists of infinitely many elements, but that there is no obvious way to write down the elements of Y explicitly like we did the set E in Example (2). Even though $[5, 10)$ is a set, we don't need to use the set brackets in this case, as interval notation has a well-established meaning which we have used in many other math courses.

(4) Now and for the remainder of the course, let the symbol $\emptyset$ denote the **empty set**, that is, the unique set which consists of no elements. Written explicitly, $\emptyset = \{\ \}$.

(5) Now and for the remainder of the course, let the symbol $\mathbb{N}$ denote the set of all natural numbers, i.e. $\mathbb{N} = \{0, 1, 2, 3, ...\}$.

(6) Now and for the remainder of the course, let the symbol $\mathbb{R}$ denote the set of all real numbers. We may think of $\mathbb{R}$ geometrically as being the collection of all the points on the number line.

(7) Let $\mathbb{R}^2$ denote the set of all **ordered pairs** of real numbers. That is, let $\mathbb{R}^2$ be the set which consists of all pairs (x, y) where x and y are both real numbers. We may think of $\mathbb{R}^2$ geometrically as the set of all points on the Cartesian coordinate plane.

If (x, y) is an element of $\mathbb{R}^2$, it will often be convenient for us to write the pair as the **column vector** $\begin{bmatrix} x \\ y \end{bmatrix}$. For our purposes the two notations will be interchangeable. It is important to note here that the order matters when we talk about pairs, so in general we have $\begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} y \\ x \end{bmatrix}$.

(8) Let $\mathbb{R}^3$ be the set of all **ordered triples** of real numbers, i.e. $\mathbb{R}^3$ is the set of all triples (x, y, z) such that x, y, and z are all real numbers. $\mathbb{R}^3$ may be visualized geometrically as the set of all points in 3-dimensional Euclidean coordinate space. We will also write elements (x, y, z) of $\mathbb{R}^3$ be using the column vector notation $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

(9) Lastly and most generally, let $n \geq 1$ be any natural number. We will let $\mathbb{R}^n$ be the set of all **ordered n-tuples** of real numbers, i.e. the set of all n-tuples $(x_1, x_2, ..., x_n)$ for which each coordinate x_i, $1 \leq i \leq n$, is a real number. We will also use the column vector notation $\begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$ in this context.

Definition 2 (Set Notation)**.** If A is a set and x is an element of A, then we write:

$$x \in A.$$

If B is a set such that every element of B is an element of A (i.e. if $x \in B$ then $x \in A$), then we call B a **subset** of A and we write:

$$B \subseteq A.$$

In order to distinguish particular subsets we wish to talk about, we will frequently use **set-builder notation**, which for convenience we will describe informally using examples, rather than give a formal definition. For an example, suppose we wish to formally describe the set E of all even positive integers (See Example 1 (2)). Then we may write

$$E = \{x \in \mathbb{N} : x \text{ is evenly divisible by } 2\}.$$

The above notation should be read as *The set of all x in $\mathbb{N}$ such that x is evenly divisible by* 2, which clearly and precisely defines our set E. For another example, we could write

$$Y = \{x \in \mathbb{R} : 5 \leq x < 10\},$$

which reads *The set of all x in $\mathbb{R}$ such that* 5 *is less than or equal to x and x is strictly less than* 10. The student should easily verify that $Y = [5, 10)$ from Example 1 (3). In general, given a set A and a precise mathematical sentence $P(x)$ about a variable x, the set-builder notation should be read as follows.

{	$x \in A$	:	$P(x)\}$
"The set of all	elements x in A	such that	sentence $P(x)$ is true for the element x.

2 Vector Spaces and Subspaces.

Definition 3. A **(real) vector space** is a nonempty set V, whose elements are called **vectors**, together with an operation $+$, called **addition**, and an operation $\cdot$, called **scalar multiplication**, which satisfy the following ten axioms:

Addition Axioms.

(1) If $\vec{u} \in V$ and $\vec{v} \in V$, then $\vec{u} + \vec{v} \in V$.
(Closure under addition.)

(2) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ for all $\vec{u}, \vec{v} \in V$.
(Commutative property of addition.)

(3) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ for all $\vec{u}, \vec{v}, \vec{w} \in V$.
(Associative property of addition.)

(4) There exists a vector $\vec{0} \in V$ which satisfies $\vec{u} + \vec{0} = \vec{u}$ for all $\vec{u} \in V$.
(Existence of an additive identity.)

(5) For every $\vec{u} \in V$, there exists a vector $-\vec{u} \in V$ such that $\vec{u} + (-\vec{u}) = \vec{0}$.
(Existence of additive inverses.)

Scalar multiplication axioms.

(6) If $\vec{u} \in V$ and $c \in \mathbb{R}$, then $c \cdot \vec{u} \in V$.
(Closure under scalar multiplication.)

(7) $c \cdot (\vec{u} + \vec{v}) = c \cdot \vec{u} + c \cdot \vec{v}$ for all $c \in \mathbb{R}$, $\vec{u}, \vec{v} \in V$.
(First distributive property of multiplication over addition.)

(8) $(c + d) \cdot \vec{u} = c \cdot \vec{u} + d \cdot \vec{u}$ for all $c, d \in \mathbb{R}$, $\vec{u} \in V$.
(Second distributive property of multiplication over addition.)

(9) $c \cdot (d \cdot \vec{u}) = (c \cdot d) \cdot \vec{u}$ for all $c, d \in \mathbb{R}$, $\vec{u} \in V$.
(Associative property of scalar multiplication.)

(10) $1 \cdot \vec{u} = \vec{u}$ for every $\vec{u} \in V$.

We use the "arrow above" notation to help differentiate vectors ($\vec{u}$, $\vec{v}$, etc.), which may or may not be real numbers, from scalars, which are always real numbers. When no confusion will arise, we will often drop the $\cdot$ symbol in scalar multiplcation and simply write $c\vec{u}$ instead of $c \cdot \vec{u}$, $c(\vec{u} + \vec{v})$ instead of $c \cdot (\vec{u} + \vec{v})$, etc.

Example 2. Let V be an arbitrary vector space.

(1) Prove that $\vec{0} + \vec{u} = \vec{u}$ for every $\vec{u} \in V$.

(2) Prove that the zero vector $\vec{0}$ is *unique*. That is, prove that if $\vec{w} \in V$ has the property that $\vec{u} + \vec{w} = \vec{u}$ for every $\vec{u} \in V$, then we must have $\vec{w} = \vec{0}$.

(3) Prove that for every $\vec{u} \in V$, the additive inverse $-\vec{u}$ is unique. That is, prove that if $\vec{w} \in V$ has the property that $\vec{u} + \vec{w} = \vec{0}$, then we must have $\vec{w} = -\vec{u}$.

Proof. (1) By Axiom (2), the commutativity of addition, we have $\vec{0} + \vec{u} = \vec{u} + \vec{0}$. Hence by Axiom (4), we have $\vec{0} + \vec{u} = \vec{u} + \vec{0} = \vec{u}$. $\square$

(2) Suppose $\vec{w} \in V$ has the property that $\vec{u} + \vec{w} = \vec{u}$ for every $\vec{u} \in V$. Then in particular, we have $\vec{0} + \vec{w} = \vec{0}$. But $\vec{0} + \vec{w} = \vec{w}$ by part (1) above; so $\vec{w} = \vec{0} + \vec{w} = \vec{0}$. □

(3) Let $\vec{u} \in V$, and suppose $\vec{w} \in V$ has the property that $\vec{u} + \vec{w} = \vec{0}$. Let $-\vec{u}$ be the additive inverse of $\vec{u}$ guaranteed by Axiom (5). Adding $-\vec{u}$ to both sides of the equality above, and applying Axioms (2) and (3) (commutativity and associativity), we get

$$\begin{aligned} -\vec{u} + (\vec{u} + \vec{w}) &= -\vec{u} + \vec{0} \\ (-\vec{u} + \vec{u}) + \vec{w} &= -\vec{u} \\ (\vec{u} + (-\vec{u})) + \vec{w} &= -\vec{u} \\ \vec{0} + \vec{w} &= -\vec{u}. \end{aligned}$$

Now by part (1) above, we have $\vec{w} = \vec{0} + \vec{w} = -\vec{u}$.

□

Example 3 (Examples of Vector Spaces)**.** (1) The real number line $\mathbb{R}$ is a vector space, where both $+$ and $\cdot$ are interpreted in the usual way. In this case the axioms are the familiar properties of real numbers which we learn in elementary school.

(2) Consider the plane $\mathbb{R}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x, y \in \mathbb{R} \right\}$. Define an operation $+$ on $\mathbb{R}^2$ by the rule

$$\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} x+z \\ y+w \end{bmatrix} \text{ for all } x, y, z, w \in \mathbb{R},$$

and a scalar multiplication $\cdot$ by the rule

$$c \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} cx \\ cy \end{bmatrix} \text{ for every } c \in \mathbb{R},\ x, y \in \mathbb{R}.$$

Then $\mathbb{R}^2$ becomes a vector space. (Verify that each of the axioms holds.)

(3) Consider the following purely geometric description. Let V be the set of all arrows in two-dimensional space, with two arrows being regarded as equal if they have the same length and point in the same direction. Define an addition $+$ on V as follows: if $\vec{u}$ and $\vec{v}$ are two arrows in V, then lay them end-to-end, so the base of $\vec{v}$ lies at the tip of $\vec{u}$. Then define $\vec{u} + \vec{v}$ to be the arrow which shares its base with $\vec{u}$ and its tip with $\vec{v}$. (A picture helps here.) Define a scalar multiplication $\cdot$ by letting $c \cdot \vec{u}$ be the arrow which points in the same direction as $\vec{u}$, but whose length is c times the length of $\vec{u}$. Is V a vector space? (What is the relationship of V with $\mathbb{R}^2$?)

(4) In general if $n \in \mathbb{N}$, $n \geq 1$, then $\mathbb{R}^n$ is a vector space, where the addition and scalar multiplication are **coordinate-wise** a la part (2) above.

(5) Let $n \in \mathbb{N}$, and let $\mathbb{P}_n$ denote the set of all polynomials of degree at most n. That is, $\mathbb{P}_n$ consists of all polynomials of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

where the coefficients $a_0, a_1, ..., a_n$ are real numbers, and x is an abstract variable. Define $+$ and $\cdot$ as follows: Suppose $c \in \mathbb{R}$, and $p, q \in \mathbb{P}_n$, so $p(x) = a_0 + a_1 x + ... + a_n x^n$ and $q(x) = b_0 + b_1 x + ... + b_n x^n$ for some coefficients $a_0, ..., a_n, b_0, ..., b_n \in \mathbb{R}$. Then

$$(p+q)(x) = p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + ... + (a_n + b_n)x^n$$

and

$$(cp)(x) = cp(x) = ca_0 + ca_1x + ... + ca_nx^n.$$

Then $\mathbb{P}_n$ is a vector space.

Definition 4. Let V be a vector space, and let $W \subseteq V$. If W is also a vector space using the same operations $+$ and $\cdot$ inherited from V, then we call W a **vector subspace**, or just **subspace**, of V.

Example 4. For each of the following, prove or disprove your answer.

(1) Let $V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x \geq 0, y \geq 0 \right\}$, so V is the first quadrant of the Cartesian plane. Is V a vector subspace of $\mathbb{R}^2$?

(2) Let V be the set of all points on the graph of $y = 5x$. Is V a vector subspace of $\mathbb{R}^2$?
(3) Let V be the set of all points on the graph of $y = 5x + 1$. Is V a vector subspace of $\mathbb{R}^2$?
(4) Is $\mathbb{R}$ a vector subspace of $\mathbb{R}^2$?
(5) Is $\{\vec{0}\}$ a vector subspace of $\mathbb{R}^2$? (This space is called the **trivial space**.)

Proof. (2) We claim that the set V of all points on the graph of $y = 5x$ is indeed a subspace of $\mathbb{R}^2$. To prove this, observe that since V is a subset of $\mathbb{R}^2$ with the same addition and scalar multiplication operations, it is immediate that Axioms (2), (3), (7), (8), (9), and (10) hold (Verify this in your brain!) So we need only check Axioms (1), (4), (5), and (6).

If $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \in V$, then by the definition of V we have $y_1 = 5x_1$ and $y_2 = 5x_2$. It follows that $y_1+y_2 = 5x_1+5x_2 = 5(x_1+x_2)$. Hence the sum $\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}$ satisfies the defining condition of V, and so $\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \in V$. So V is closed under addition and Axiom (1) is satisfied.

Check that the additive identity $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in V, so Axiom (3) is satisfied.

If $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \in V$ and $c \in \mathbb{R}$, then $c\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cy_1 \end{bmatrix} \in V$ since $cy_1 = c(5x_1) = 5(cx_1)$, so V is closed under scalar multiplication. Hence Axiom (6) is satisfied. Moreover, if we take $c = -1$ in the previous equalities, we see that each vector $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \in V$ has an additive inverse $\begin{bmatrix} -x_1 \\ -y_1 \end{bmatrix} \in V$. So Axiom (5) is satisfied. Thus V meets all the criteria to be a vector space, as we claimed. □

(3) On the other hand, if we take V to be the set of all points on the graph of $y = 5x + 1$, then V is *not* a vector subspace of $\mathbb{R}^2$. To see this, it suffices to check, for instance, that V fails Axiom (1), i.e. V is not closed under addition.

To show that V is not closed under addition, it suffices to exhibit two vectors in V whose sum is not in V. So let $\vec{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and let $\vec{v} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. Both $\vec{u}$ and $\vec{v}$ are in V. (Why?) But their sum $\vec{u} + \vec{v} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$ is *not* a solution of the equation $y = 5x + 1$, and hence not in V. So V fails to satisfy Axiom (1), and cannot be a vector space.

□

We will not prove the following fact, but the reader should think about why it must be true.

Fact 1. *Let V be a vector space, and let $W \subseteq V$. Then W, together with the addition and scalar multiplication inherited from V, satisfies Axioms (2), (3), (7), (8), (9), (10) in the definition of a vector space.*

Theorem 1. *Let V be a vector space and let $W \subseteq V$. Then W is a subspace of V if and only if the following three properties hold:*

(1) $\vec{0} \in W$.
(2) W is closed under addition. That is, for each $\vec{u}, \vec{v} \in W$, we have $\vec{u} + \vec{v} \in W$.
(3) W is closed under scalar multiplication. That is, for each $c \in \mathbb{R}$ and $\vec{u} \in W$, we have $c\vec{u} \in W$.

Proof. One direction of the proof is trivial: if W is a vector subspace of V, then W satisfies the three conditions above because it satisfies Axioms (4), (1), and (6) respectively in the definition of a vector space.

Conversely, suppose $W \subseteq V$, and W satisfies the three conditions above. The three conditions imply that W satisfies Axioms (4), (1), and (6), respectively, while our previous fact implies that W also satisfies Axioms (2), (3), (7), (8), (9), and (10). So the only axiom left to check is (5).

To see that (5) is satisfied, let $\vec{u} \in W$. Since $W \subseteq V$, the vector $\vec{u}$ is in V, and hence has an additive inverse $-\vec{u} \in V$. We must show that in fact $-\vec{u} \in W$. Note that since W is closed under scalar multiplication, the vector $-1 \cdot \vec{u}$ is in W. But on our homework problem #1(d), we show that in fact $-1 \cdot \vec{u} = -\vec{u}$. So $-\vec{u} = -1 \cdot \vec{u} \in V$ as we hoped, and the proof is complete. □

Example 5. What do the vector subspaces of $\mathbb{R}^2$ look like? What about $\mathbb{R}^3$?

3 Linear Combinations and Spanning Sets.

Definition 5. Let V be a vector space. Let $\vec{v_1}, \vec{v_2}, ..., \vec{v_n} \in V$, and let $c_1, c_2, ..., c_n \in \mathbb{R}$. We say that a vector $\vec{u}$ defined by

$$\vec{u} = c_1\vec{v_1} + c_2\vec{v_2} + ... + c_n\vec{v_n}$$

is a **linear combination** of the vectors $\vec{v_1}, ..., \vec{v_n}$ with **weights** $c_1, ..., c_n$. Notice that since addition is associative in any vector space V, we may omit parentheses from the sum above. Also notice that the zero vector $\vec{0}$ is always a linear combination of any collection of vectors $\vec{v_1}, ..., \vec{v_n}$, since we may always take $c_1 = c_2 = ... = c_n = 0$.

Example 6. Let $\vec{v_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and let $\vec{v_2} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

(1) Which points in $\mathbb{R}^2$ are linear combinations of v_1 and v_2, using *integer* weights?
(2) Which points in $\mathbb{R}^2$ are linear combinations of v_1 and v_2, using any weights?

Example 7. Let $\vec{a_1} = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$ and $\vec{a_2} = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$ be vectors in $\mathbb{R}^3$.

(1) Let $\vec{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$. May $\vec{b}$ be written as a linear combination of $\vec{a_1}$ and $\vec{a_2}$?
(2) Let $\vec{b} = \begin{bmatrix} 6 \\ 3 \\ -5 \end{bmatrix}$. May $\vec{b}$ be written as a linear combination of $\vec{a_1}$ and $\vec{a_2}$?

Partial Solution. (1) We wish to answer the question: Do there exist real numbers c_1 and c_2 for which

$$c_1\vec{a_1} + c_2\vec{a_2} = \vec{b}?$$

Written differently, are there $c_1, c_2 \in \mathbb{R}$ for which

$$c_1 \cdot \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + c_2 \cdot \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} c_1 + 2c_2 \\ -2c_1 + 5c_2 \\ -5c_1 + 6c_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}?$$

Recall that vectors in $\mathbb{R}^3$ are equal if and only if each pair of entries are equal. So to identify a c_1 and c_2 which make the above equality true, we wish to solve the following system of equations in two variables:

$$\begin{aligned} c_1 + 2c_2 &= 7 \\ -2c_1 + 5c_2 &= 4 \\ -5c_1 + 6c_2 &= -3 \end{aligned}$$

This can be done manually using elementary algebra techniques. We should get $c_1 = 3$ and $c_2 = 2$. Since a solution exists, $\vec{b} = 3\vec{a_1} + 2\vec{a_2}$ is indeed a linear combination of $\vec{a_1}$ and $\vec{a_2}$.

□

Definition 6. Let V be a vector space and $\vec{v_1}, ..., \vec{v_n} \in V$. We denote the set of all possible linear combinations of $\vec{v_1}, ..., \vec{v_n}$ by $\text{Span}\{\vec{v_1}, ..., \vec{v_n}\}$, and we call this set the subset of V **spanned** by $\vec{v_1}, ..., \vec{v_n}$. We also call $\text{Span}\{\vec{v_1}, ..., \vec{v_n}\}$ the subset of V **generated** by $\vec{v_1}, ..., \vec{v_n}$.

Example 8. (1) Find Span $\left\{ \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right\}$ in $\mathbb{R}^2$.

(2) Find Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ in $\mathbb{R}^3$.

Theorem 2. *Let V be a vector space and let $W \subseteq V$. Then W is a subspace of V if and only if W is closed under linear combinations, i.e., for every $\vec{v_1}, ..., \vec{v_n} \in W$ and every $c_1, ..., c_n \in \mathbb{R}$, we have $c_1\vec{v_1} + ... + c_n\vec{v_n} \in W$.*

Corollary 1. *Let V be a vector space and $\vec{v_1}, ..., \vec{v_n} \in V$. Then $\text{Span}\{\vec{v_1}, ..., \vec{v_n}\}$ is a subspace of V.*

4 Matrix Row Reductions and Echelon Forms.

As our previous few examples should indicate, our techniques for solving systems of linear equations are extremely relevant to our understanding of linear combinations and spanning sets. The student should recall the basics of solving systems from a previous math course, but in this section we will develop a new notationally convenient method for finding solutions to such systems.

Definition 7. Recall that a **linear equation** in n variables is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

where $a_1, ..., a_n, b \in \mathbb{R}$ and $x_1, ..., x_n$ are variables. A **system** of linear equations is any finite collection of linear equations involving the same variables $x_1, ..., x_n$. A **solution** to the system is an n-tuple $(s_1, s_2, ..., s_n) \in \mathbb{R}^n$, that makes each equation in the system true if we substitute $s_1, ..., s_n$ for $x_1, ..., x_n$ respectively.

The set of all possible solutions to a system is called the **solution set**. Two systems are called **equivalent** if they have the same solution set. It is only possible for a system of linear equations to have no solutions, exactly one solution, or infinitely many solutions. (Geometrically one may think of "parallel lines," "intersecting lines," and "coincident lines," respectively.

A system of linear equations is called **consistent** if it has at least one solution; if the system's solution set is $\emptyset$, then the system is **inconsistent**.

Example 9. Solve the following system of three equations in three variables.

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_2 & = & 8 \\ -4x_1 & + & 5x_2 & + & 9x_3 & = & -9 \end{array}$$

Solution. Our goal here will be to solve the system using the elimination technique, but with a stripped down notation which preserves only the necessary information and makes computations by hand go quite a bit faster.

First let us introduce a little terminology. Given the system we are working with, the **matrix of coefficients** is the 3x3 matrix below.

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$$

The **augmented matrix** of the system is the 3x4 matrix below.

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right]$$

Note that the straight black line between the third and fourth columns in the matrix above is not necessary. Its sole purpose is to remind us where the "equals sign" was in our original system, and may be omitted if the student prefers.

To solve our system of equations, we will perform operations on the augmented matrix which "encode" the process of elimination. For instance, if we were using elimination on the given system, we might add 4 times the first equation to the third equation, in order to eliminate the x_1 variable from the third equation. Let's do the analogous operation to our matrix instead.

Add 4 *times the top row to the bottom row:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right]$$

Notice that the new matrix we obtained above can also be interpreted as the augmented matrix of a system of linear equations, which is the new system we would have obtained by just using the elimination technique. In particular, the first augmented matrix and the new augmented matrix represent systems which are **equivalent** to one another. (However, the two matrices are obviously not **equal** to one another, which is why we will stick to the arrow $\mapsto$ notation instead of using equals signs!)

Now we will continue the process, replacing our augmented matrix with a new, simpler augmented matrix, in such a way that the linear systems the matrices represent are all equivalent to one another.

Multiply the second row by $\frac{1}{2}$*:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & -3 & 13 & -9 \end{array}\right]$$

Add 3 *times the second row to the third row:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & -3 & 13 & -9 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Add 4 *times the third row to the second row:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Add -1 *times the third row to the first row:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & -2 & 0 & -3 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Add 2 *times the second row to the first row:*

$$\left[\begin{array}{rrr|r} 1 & -2 & 0 & -3 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Now let's stop and examine what we've done. The last augmented matrix above represents the following system:

$$\begin{array}{ccccc} x_1 & & & = & 29 \\ & x_2 & & = & 16 \\ & & x_3 & = & 3 \end{array}$$

So the system is solved, and has a unique solution $(29, 16, 3)$. □

Definition 8. Given an augmented matrix, the three **elementary row operations** are the following.

(1) (Replacement) Replace one row by the sum of itself and a multiple of another row.
(2) (Interchange) Interchange two rows.
(3) (Scaling) Multiply all entries in a row by a nonzero constant.

These are the "legal moves" when solving systems of equations using augmented matrices. Two matrices are called **row equivalent** if one can be transformed into the other using a finite sequence of elementary row operations.

Fact 2. *If two augmented matrices are row equivalent, then the two linear systems they represent are equivalent, i.e. they have the same solution set.*

Definition 9. A matrix is said to be in **echelon form**, or **row echelon form**, if it has the following three properties:

(1) All rows which have nonzero entries are above all rows which have only zero entries.
(2) Each leading nonzero entry of a row is in a column to the right of the leading nonzero entry of the row above it.
(3) All entries in a column below a leading nonzero entry are zeros.

A matrix is said to be in **reduced echelon form** or **reduced row echelon form** if it is in echelon form, and also satisfies the following two properties:

(4) Each nonzero leading entry is 1.
(5) Each leading 1 is the only nonzero entry in its column.

Every matrix may be transformed, via elementary row operations, into a matrix in reduced row echelon form. This process is called **row reduction**.

Example 10. Use augmented matrices and row reductions to find solution sets for the following systems of equations.

$$(1)\quad \begin{array}{rcrcrcr} 2x_1 & & & - & 6x_3 & = & -8 \\ & & x_2 & + & 2x_3 & = & 3 \\ 3x_1 & + & 6x_2 & - & 2x_3 & = & -4 \end{array}$$

$$(2)\quad \begin{array}{rcrcrcr} x_1 & - & 5x_2 & + & 4x_3 & = & -3 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \\ -2x_1 & + & x_2 & + & 7x_3 & = & -1 \end{array}$$

5 Functions and Function Notation

Definition 10. Let X and Y be sets. A **function** f from X to Y is just a rule by which we associate to each element $x \in X$, an element $f(x) \in Y$. The input set X is called the **domain** of f, while the set Y of possible outputs is called the **codomain** of f. To denote the domain and codomain when we define a function, we write

$$f : X \to Y.$$

We regard the terms **map**, **mapping**, and **transformation** as all being synonymous with function.

Example 11. (1) For the familiar quadratic function $f(x) = x^2$, we would write $f : \mathbb{R} \to \mathbb{R}$, since f takes real numbers for input and returns real numbers for output. Notice that the codomain $\mathbb{R}$ here is distinct from the **range**, i.e. the set of all actual outputs of the function, which the reader should know is just the set $[0, \infty)$. This is an ok part of the notation.
(2) On the other hand, for the familiar hyperbolic function $f(x) = \frac{1}{x}$, we would NOT write $f : \mathbb{R} \to \mathbb{R}$; this is because $0 \in \mathbb{R}$ but 0 is not part of the domain of f.

6 Linear Transformations

Definition 11. Let V and W be vector spaces. A function $T : V \to W$ is a **linear transformation** if

(1) $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$ for every $\vec{v}, \vec{w} \in V$.

(2) $T(c\vec{v}) = cT(\vec{v})$ for every $\vec{v} \in V$ and $c \in \mathbb{R}$.

Note that the transformations which are linear are those which **respect the addition and scalar multiplication operations** of the vector spaces involved.

Example 12. Are the following maps linear transformations?

(1) $T : \mathbb{R} \to \mathbb{R}$, $T(x) = \begin{bmatrix} x \\ 3x \end{bmatrix}$.

(2) $T : \mathbb{R} \to \mathbb{R}^2$, $T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$.

(3) $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(\vec{v}) = A\vec{v}$, where $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$.

Fact 3. *Every **matrix transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

Fact 4. *If $T : V \to W$ is a linear transformation, then T **preserves linear combinations**. In other words, for every collection of vectors $\vec{v_1}, ..., \vec{v_n} \in V$ and every choice of weights $c_1, ..., c_n \in \mathbb{R}$, we have*

$$T(c_1\vec{v_1} + ... + c_n\vec{v_n}) = c_1T(\vec{v_1}) + ... + c_nT(\vec{v_n}).$$

Definition 12. Let $V = \mathbb{R}^n$ be n-dimensional Euclidean space. Define vectors $\vec{e_1}, ..., \vec{e_n} \in V$ by:

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ ... \\ 0 \\ 0 \end{bmatrix};\ \vec{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ ... \\ 0 \\ 0 \end{bmatrix};\ ...;\ \vec{e_n} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \\ 1 \end{bmatrix}.$$

These vectors $\vec{e_1}, ..., \vec{e_n}$ are called the **standard basis** for $\mathbb{R}^n$. Note that any vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix}$ may be written as a linear combination of the standard basis vectors in an obvious way: $\vec{v} = v_1\vec{e_1} + v_2\vec{e_2} + ... + v_n\vec{e_n}$.

Example 13. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation such that $T(\vec{e_1}) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix}$ and $T(\vec{e_2}) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$. Find an explicit formula for T.

Theorem 3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N})$ be a linear transformation. Then there exists a unique $m \times n$ matrix A such that $T(\vec{v}) = A\vec{v}$ for every $\vec{v} \in \mathbb{R}^n$. In fact, we have*

$$A = \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & ... & T(\vec{e_n}) \end{bmatrix}.$$

Proof. Notice that for any $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} \in \mathbb{R}^n$, we have $\vec{v} = v_1\vec{e_1} + ... + v_n\vec{e_n}$. Since T is a linear transformation, it respects linear combinations, and hence

$$\begin{aligned} T(\vec{v}) &= v_1T(\vec{e_1}) + ... + v_nT(\vec{e_n}) \\ &= \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & ... & T(\vec{e_n}) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} \\ &= A\vec{v}, \end{aligned}$$

where A is as we defined in the statement of the theorem. □

Example 14. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which scales up every vector by a factor of 3, i.e. $T(\vec{v}) = 3\vec{v}$ for every $\vec{v} \in \mathbb{R}^2$. Find a matrix representation for T.

Example 15. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which rotates the plane about the origin at some fixed angle θ. Is T a linear transformation? If so, find its matrix representation.

7 The Null Space of a Linear Transformation

Definition 13. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Define the following set:

$$\operatorname{Nul} T = \{\vec{v} \in V : T(\vec{v}) = \vec{0}\}.$$

Then this set $\operatorname{Nul} T$ is called the **null space** or **kernel** of the map T.

Example 16. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the matrix transformation $T(\vec{v}) = A\vec{v}$, where $A = \begin{bmatrix} 1 & -1 & 5 \\ 2 & 0 & 7 \\ -3 & -5 & -3 \end{bmatrix}$.

(1) Let $\vec{u} = \begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix}$. Is $\vec{u} \in \operatorname{Nul} T$?

(2) Let $\vec{w} = \begin{bmatrix} -5 \\ -5 \\ 0 \end{bmatrix}$. Is $\vec{w} \in \operatorname{Nul} T$?

Theorem 4. *Let V, W be vector spaces and $T : V \to W$ a linear transformation. Then $\operatorname{Nul} T$ is a vector subspace of V.*

Proof. We will show that $\operatorname{Nul} T$ is closed under taking linear combinations, and hence the result will follow from Theorem 2.

Let $\vec{v_1}, ..., \vec{v_n} \in \operatorname{Nul} T$ and $c_1, ..., c_n \in \mathbb{R}$ be arbitrary. We must show that $c_1\vec{v_1} + ... + c_n\vec{v_n} \in \operatorname{Nul} T$, i.e. that $T(c_1\vec{v_1} + ... + c_n\vec{v_n}) = \vec{0}$. To see this, simply observe that since T is a linear transformation, we have

$$T(c_1\vec{v_1} + ... + c_n\vec{v_n}) = c_1T(\vec{v_1} + ... + c_nT(\vec{v_n}).$$

But since $\vec{v_1}, ..., \vec{v_n} \in \operatorname{Nul} T$, we have $T(\vec{v_1}) = ... = T(\vec{v_n}) = \vec{0}$. So in fact we have

$$T(c_1\vec{v_1} + ... + c_n\vec{v_n}) = c_1\vec{0} + ... + c_n\vec{0} = \vec{0}.$$

It follows that $c_1\vec{v_1} + ... + c_n\vec{v_n} \in \operatorname{Nul} T$, and $\operatorname{Nul} T$ is closed under linear combinations. This completes the proof. □

Example 17. For the following matrix transformations, give an explicit description of $\operatorname{Nul} T$ by finding a spanning set.

(1) $T : \mathbb{R}^4 \to \mathbb{R}^2$, $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 4 & 0 \\ 0 & 1 & 3 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ for every $x_1, x_2, x_3, x_4 \in \mathbb{R}$.

(2) $T : \mathbb{R}^5 \to \mathbb{R}^3$, $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} 1 & -4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -5 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$ for every $x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}$.

Definition 14. Let V, W be sets. A mapping $T : V \to W$ is called **one-to-one** if the following holds: for every $\vec{v}, \vec{w} \in V$, if $\vec{v} \neq \vec{w}$, then $T(\vec{v}) \neq T(\vec{w})$.

Equivalently, T is one-to-one if the following holds: for every $\vec{v}, \vec{w} \in V$, if $T(\vec{v}) = T(\vec{w})$, then $\vec{v} = \vec{w}$).

A map is one-to-one only if it sends distinct elements in V to distinct elements in W, i.e. no two elements in V are mapped to the same place in W.

Example 18. Are the following linear transformations one-to-one?

(1) $T : \mathbb{R} \to \mathbb{R}^2$, $T(x) = \begin{bmatrix} x \\ 3x \end{bmatrix}$.

(2) $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Theorem 5. *A linear transformation $T : V \to W$ is one-to-one if and only if* $\operatorname{Nul} T = \{\vec{0}\}$.

Proof. ($\Rightarrow$) First suppose T is one-to-one, and let $\vec{v} \in \operatorname{Nul} T$. We will show $\vec{v} = \vec{0}$. To see this, note that $T(\vec{v}) = \vec{0}$ since $\vec{v} \in \operatorname{Nul} T$. But we also have $\vec{0} \in \operatorname{Nul} T$ since $\operatorname{Nul} T$ is a subspace of V, so $T(\vec{0}) = \vec{0} = T(\vec{v})$. Since T is one-to-one, we must have $\vec{0} = \vec{v}$, and hence $\operatorname{Nul} T$ is the trivial subspace $\operatorname{Nul} T = \{\vec{0}\}$.

($\Leftarrow$) Conversely, suppose T is not one-to-one; we will show $\operatorname{Nul} T$ is non-trivial. Since T is not one-to-one, there exist two distinct vectors $\vec{v}, \vec{w} \in V$, $\vec{v} \neq \vec{w}$, such that $T(\vec{v}) = T(\vec{w})$. Set $\vec{u} = \vec{v} - \vec{w}$. Since $\vec{v} \neq \vec{w}$ and additive inverses are unique, we have $\vec{u} \neq \vec{0}$, and we also have

$$T(\vec{u}) = T(\vec{v} - \vec{w}) = T(\vec{v}) - T(\vec{w}) = T(\vec{v}) - T(\vec{v}) = \vec{0}.$$

So $\vec{u} \in \operatorname{Nul} T$. Since $\vec{u}$ is not the zero vector, $\operatorname{Nul} T \neq \{\vec{0}\}$. ☐

Definition 15. Let V, W be sets. A mapping $T : V \to W$ is called **onto** if the following holds: for every $\vec{w} \in W$, there exists a vector $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$.

Example 19. Are the following linear transformations onto?

(1) $T : \mathbb{R} \to \mathbb{R}^2$, $T(x) = \begin{bmatrix} x \\ 3x \end{bmatrix}$.

(2) $T : \mathbb{R}^2 \to \mathbb{R}$, $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_2 - x_1$.

(3) $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_2 \\ 2x_1 + x_2 \end{bmatrix}$.

Definition 16. Let V, W be vector spaces and $T : V \to W$ a linear transformation. If T is both one-to-one and onto, then T is called an **isomorphism**. In this case the domain V and codomain W are called **isomorphic as vector spaces**, or just **isomorphic**. It means that V and W are indistinguishable from one another in terms of their vector space structure.

Example 20. Prove that the mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$, where T rotates the plane by a fixed angle θ, is an isomorphism.

Example 21. Let W be the graph of the line $y = 3x$, a vector subspace of $\mathbb{R}^2$. Prove that $\mathbb{R}$ is isomorphic to W.

8 The Range Space of a Linear Transformation and the Column Space of a Matrix

Definition 17. Let V, W be vector spaces and $T : V \to W$ a linear transformation. Define the following set:

$$\operatorname{Ran} T = \{\vec{w} \in W : \text{there exists a vector } \vec{v} \in V \text{ such that } T(\vec{v}) = \vec{w}\}.$$

Then $\operatorname{Ran} T$ is called the **range** of T. (This definition should coincide with the student's knowledge of the range of a function from previous courses.)

Theorem 6. *Let V, W be vector spaces and $T : V \to W$ a linear transformation. Then $\operatorname{Ran} T$ is a vector subspace of W.*

Proof. Again we will show that $\operatorname{Ran} T$ is closed under linear combinations, and appeal to Theorem 2.

To that end, let $\vec{w_1}, ..., \vec{w_n} \in \operatorname{Ran} T$ and let $c_1, ..., c_n \in \mathbb{R}$ all be arbitrary. We must show $c_1\vec{w_1} + ... + c_n\vec{w_n} \in \operatorname{Ran} T$. To see this, note that since $\vec{w_1}, ..., \vec{w_n} \in \operatorname{Ran} T$, there exist vectors $\vec{v_1}, ..., \vec{v_n} \in V$ such that $T(\vec{v_1}) = \vec{w_1}, ..., T(\vec{v_n}) = \vec{w_n}$. Set

$$\vec{u} = c_1\vec{v_1} + ... + c_n\vec{v_n}.$$

Since V is a vector space, V is closed under taking linear combinations and hence $\vec{u} \in V$. Moreover, we have

$$\begin{aligned} T(\vec{u}) &= T(c_1\vec{v_1} + ... + c_n\vec{v_n}) \\ &= c_1T(\vec{v_1}) + ... + c_nT(\vec{v_n}) \\ &= c_1\vec{w_1} + ... + c_n\vec{w_n}. \end{aligned}$$

So the vector $c_1\vec{w_1} + ... + c_n\vec{w_n}$ is the image of $\vec{u}$ under T, and hence $c_1\vec{w_1} + ... + c_n\vec{w_n} \in \operatorname{Ran} T$. This shows $\operatorname{Ran} T$ is closed under linear combinations and hence a vector subspace of W. □

Theorem 7. *Let V, W be vector spaces and $T : V \to W$ a linear transformation. Then T is onto if and only if $\operatorname{Ran} T = W$.*

Proof. Obvious if you think about it! □

Corollary 2. *Let V, W be vector spaces and $T : V \to W$ a linear transformation. Then the following statements are all equivalent:*

(1) T is an isomorphism.

(2) T is one-to-one and onto.

(3) $\operatorname{Nul} T = \{\vec{0}\}$ and $\operatorname{Ran} T = W$.

Definition 18. Let $m, n \in \mathbb{N}$, and let A be an $m \times n$ matrix (so A induces a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$). Write

$$A = \begin{bmatrix} \vec{w_1} & ... & \vec{w_n} \end{bmatrix},$$

where each of $\vec{w_1}, ..., \vec{w_n}$ is a column vector in $\mathbb{R}^m$.

Define the following set:

$$\operatorname{Col} A = \operatorname{Span}\{\vec{w_1}, ..., \vec{w_n}\}.$$

Then $\operatorname{Col} A$ is called the **column space** of the matrix A. $\operatorname{Col} A$ is exactly the set of all vectors in W which may be written as a linear combination of the columns of the matrix A. Of course $\operatorname{Col} A$ is also a vector subspace of $\mathbb{R}^m$, by Corollary 1.

Example 22. Let $A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 0 & 0 \end{bmatrix}$.

(1) Determine whether or not the vector $\begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}$ is in $\mathrm{Col}\, A$.

(2) What does $\mathrm{Col}\, A$ look like in $\mathbb{R}^3$?

Example 23. Let $W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Find a matrix A so that $W = \mathrm{Col}\, A$.

Theorem 8. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N})$ be a linear transformation. Let A be the $m \times n$ matrix representation of A guaranteed by Theorem 3. Then* $\mathrm{Ran}\, T = \mathrm{Col}\, A$.

Proof. Recall from Theorem 3 that the matrix A is given by

$$A = \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & ... & T(\vec{e_n}) \end{bmatrix},$$

where $\vec{e_1}, ..., \vec{e_n}$ are the standard basis vectors in $\mathbb{R}^n$. So its columns are just $T(\vec{e_1}), ..., T(\vec{e_n})$.

The first thing we will show is that $\mathrm{Ran}\, T \subseteq \mathrm{Col}\, A$. So suppose $\vec{w} \in \mathrm{Ran}\, T$. Then there exists some vector $\vec{v} \in \mathbb{R}^n$ for which $T(\vec{v}) = \vec{w}$. Write $\vec{v} = \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix} = v_1\vec{e_1} + ... + v_n\vec{e_n}$ for some real numbers $v_1, ..., v_n \in \mathbb{R}$. Then, since T is a linear transformation, we have

$$\begin{aligned} \vec{w} &= T(\vec{v}) \\ &= T(v_1\vec{e_1} + ... + v_n\vec{e_n}) \\ &= v_1 T(\vec{e_1} + ... + v_n T(\vec{e_n}). \end{aligned}$$

So the above equality displays $\vec{w}$ as a linear combination of the columns of A, with weights $v_1, ..., v_n$. This proves $\vec{w} \in \mathrm{Col}\, A$. Since $\vec{w}$ was taken arbitrarily out of $\mathrm{Ran}\, T$, we must have $\mathrm{Ran}\, T \subseteq \mathrm{Col}\, A$.

The next thing we will show is that $\mathrm{Col}\, A \subseteq \mathrm{Ran}\, T$. So let $\vec{w} \in \mathrm{Col}\, A$. Then $\vec{w}$ can be written as a linear combination of the columns of A, so there exist some weights $c_1, ..., c_n \in \mathbb{R}$ for which

$$\vec{w} = c_1 T(\vec{e_1}) + ... + c_n T(\vec{e_n}).$$

Set $\vec{v} = c_1\vec{e_1} + ... + c_n\vec{e_n}$. So $\vec{v} \in \mathbb{R}^n$, and we claim that $T(\vec{v}) = \vec{w}$. To see this, just compute (again using the fact that T is linear):

$$\begin{aligned} T(\vec{v}) &= T(c_1\vec{e_1} + ...c_n\vec{e_n}) \\ &= c_1 T(\vec{e_1}) + ... + c_n T(\vec{e_n}) \\ &= \vec{w}. \end{aligned}$$

This shows $\vec{w} \in \operatorname{Ran} T$, and again since $\vec{w}$ was taken arbitrarily out of $\operatorname{Col} A$, we have shown that $\operatorname{Col} A \subseteq \operatorname{Ran} T$.

Since $\operatorname{Ran} T \subseteq \operatorname{Col} A$ and $\operatorname{Col} A \subseteq \operatorname{Ran} T$, we must have $\operatorname{Ran} T = \operatorname{Col} A$. This completes the proof. $\square$

Example 24. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by $T(\vec{e_1}) = -2\vec{e_1} + 4\vec{e_2}$ and $T(\vec{e_2}) = -\vec{e_1} + 2\vec{e_2}$. Let $\vec{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Is $\vec{w} \in \operatorname{Ran} T$?

9 Linear Independence and Bases.

Definition 19. Let V be a vector space and let $\vec{v_1}, ..., \vec{v_n} \in V$. Consider the equation:

$$c_1\vec{v_1} + ... + c_n\vec{v_n} = \vec{0}$$

where $c_1, ..., c_n$ are interpreted as real variables. Notice that $c_1 = ... = c_n = 0$ is always a solution to the equation, which is called the **trivial solution**, but that there may be others depending on our choice of $\vec{v_1}, ..., \vec{v_n}$.

If there exists a nonzero solution $(c_1, ..., c_n)$ to the equation, i.e. a solution where $c_k \neq 0$ for at least one c_k $(1 \leq k \leq n)$ then the vectors $\vec{v_1}, ..., \vec{v_n}$ are called **linearly dependent**. Otherwise if the trivial solution is the only solution, then the vectors $\vec{v_1}, ..., \vec{v_n}$ are called **linearly independent**.

Example 25. Let $\vec{v_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\vec{v_2} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\vec{v_3} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

(1) Are the vectors $\vec{v_1}$, $\vec{v_2}$, $\vec{v_3}$ linearly independent?

(2) If possible, find a **dependence relation**, i.e. a non-trivial linear combination of $\vec{v_1}, \vec{v_2}, \vec{v_3}$ which sums to $\vec{0}$.

Fact 5. *Let $\vec{v_1}, ..., \vec{v_n}$ be column vectors in $\mathbb{R}^m$. Define an $m \times n$ matrix A by:*

$$A = \begin{bmatrix} \vec{v_1} & ... & \vec{v_n} \end{bmatrix},$$

and define a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by the rule $T(\vec{v}) = A\vec{v}$ for every $\vec{v} \in \mathbb{R}^n$. Then the following statements are all equivalent:

(1) The vectors $\vec{v_1}, ..., \vec{v_n}$ are linearly independent.

(2) $\operatorname{Nul} T = \{\vec{0}\}$.

(3) T is one-to-one.

Example 26. Let $\vec{v_1} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$. Are $\vec{v_1}$ and $\vec{v_2}$ linearly independent?

Fact 6. *Let V be a vector space and $\vec{v_1}, ..., \vec{v_n} \in V$. Then $\vec{v_1}, ..., \vec{v_n}$ are linearly dependent if and only if there exists some $k \in \{1, ..., n\}$ such that $\vec{v_k} \in \operatorname{Span}\{\vec{v_1}, ..., \vec{v_{k-1}}, \vec{v_{k+1}}, ..., \vec{v_n}\}$, i.e. $\vec{v_k}$ can be written as a linear combination of the other vectors.*

Example 27. Let $\vec{v_1} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathbb{R}^2$.

(1) Describe all vectors $\vec{v_2} \in \mathbb{R}^2$ for which $\vec{v_1}, \vec{v_2}$ are linearly independent.

(2) Describe all pairs of vectors $\vec{v_2}, \vec{v_3} \in \mathbb{R}^2$ for which $\vec{v_1}, \vec{v_2}, \vec{v_2}$ are linearly independent.

Example 28. Let $\vec{v_1} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}$. Describe all vectors $\vec{v_3} \in \mathbb{R}^3$ for which $\vec{v_1}, \vec{v_2}, \vec{v_3}$ are linearly independent.

Definition 20. Let V be a vector space and let $\vec{b_1}, ..., \vec{b_n} \in V$. The set $\{\vec{b_1}, ..., \vec{b_n}\} \subseteq V$ is called a **basis** for V if

(1) $\vec{b_1}, ..., \vec{b_n}$ are linearly independent, and

(2) $V = \text{Span}\{\vec{b_1}, ..., \vec{b_n}\}$.

Example 29. Let $\vec{v_1} = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$, $\vec{v_2} = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}$, and $\vec{v_3} = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$. Is $\{\vec{v_1}, \vec{v_2}, \vec{v_3}\}$ a basis for $\mathbb{R}^3$?

Example 30. Let $\vec{v_1} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$.

(1) Is $\{\vec{v_1}, \vec{v_2}\}$ a basis for $\mathbb{R}^2$?

(2) What if $\vec{v_2} = \begin{bmatrix} -10 \\ -15 \end{bmatrix}$?

Fact 7. *The standard basis $\{\vec{e_1}, ..., \vec{e_n}\}$ is a basis for $\mathbb{R}^n$.*

Fact 8. *The set $\{1, x, x^2, ..., x^n\}$ is a basis for $\mathbb{P}_n$.*

Fact 9. *Let V be any vector space and let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for V.*

(1) If you remove any one element $\vec{b_k}$ from $\mathcal{B}$, then the resulting set no longer spans V. This is because $\vec{b_k}$ is linearly independent from the other vectors in $\mathcal{B}$, and hence cannot be written as a linear combination of them by Fact 6. In this sense a basis is a "minimal spanning set."

(2) If you add any one element $\vec{v}$, not already in $\mathcal{B}$, to $\mathcal{B}$, then the result set is no longer linearly independent. This is because $\mathcal{B}$ spans V, and hence there are no vectors in V which are independent from those in $\mathcal{B}$ by Fact 6. In this sense a basis is a "maximal linearly independent set."

Theorem 9 (Unique Representation Theorem)**.** *Let V be a vector space and let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for V. Then every vector $\vec{v} \in V$ may be written as a linear combination $\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n}$ in one and only one way, i.e. $\vec{v}$ has a unique representation with respect to $\mathcal{B}$.*

Proof. Since $\{\vec{b_1}, ..., \vec{b_n}\}$ is a basis for V, in particular the set spans V, so we have $\vec{v} \in \text{Span}\{\vec{b_1}, ..., \vec{b_n}\}$. It follows that there exist some scalars $c_1, ..., c_n \in \mathbb{R}^n$ for which

$$\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n}.$$

So we need only check that the representation above is unique. To that end, suppose $d_1, ..., d_n \in \mathbb{R}$ is another set of scalars for which

$$\vec{v} = d_1\vec{b_1} + ... + d_n\vec{b_n}.$$

We will show that in fact $d_1 = c_1$, ..., $d_n = c_n$, and hence there is really only one choice of scalars to begin with. To see this, observe the following equalities:

$$\begin{aligned}(d_1 - c_1)\vec{b_1} + ... + (d_n - c_n)\vec{b_n} &= [d_1\vec{b_1} + ... + d_n\vec{b_n}] - [c_1\vec{b_1} + ... + c_n\vec{b_n}] \\ &= \vec{v} - \vec{v} \\ &= \vec{0}.\end{aligned}$$

So we have written $\vec{0}$ as a linear combination of $\vec{b_1}, ..., \vec{b_n}$. But the collection is a basis and hence linearly independent, so the coefficients must all be zero, i.e. $d_1 - c_1 = ... = d_n - c_n = 0$. The conclusion now follows. □

10 Dimension

Theorem 10 (Steinitz Exchange Lemma). *Let V be a vector space and let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for V. Let $\vec{v} \in V$. By Theorem 9 there are unique scalars $c_1, ..., c_n$ for which*

$$\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n}.$$

If there is $k \in \{1, ..., n\}$ for which $c_k \neq 0$, then exchanging $\vec{v}$ for $\vec{b_k}$ yields another basis for the space, i.e. the collection $\{\vec{b_1}, ..., \vec{b_{k-1}}, \vec{v}, \vec{b_{k+1}}, ..., \vec{b_n}\}$ is a basis for V.

Proof. Call the new collection $\hat{\mathcal{B}} = \{\vec{b_1}, ..., \vec{b_{k-1}}, \vec{v}, \vec{b_{k+1}}, ..., \vec{b_n}\}$. We must show that $\hat{\mathcal{B}}$ is both linearly independent and that it spans V.

First we check that $\hat{\mathcal{B}}$ is linearly independent. Consider the equation

$$d_1\vec{b_1} + ... + d_{k-1}\vec{b_{k-1}} + d_k\vec{v} + d_{k+1}\vec{b_{k+1}} + ... + d_n\vec{b_n} = \vec{0}.$$

Let's substitute the representation for $\vec{v}$ in the above:

$$d_1\vec{b_1} + ... + d_{k-1}\vec{b_{k-1}} + d_k(c_1\vec{b_1} + ... + c_n\vec{b_n}) + d_{k+1}\vec{b_{k+1}} + ... + d_n\vec{b_n} = \vec{0}.$$

Now collecting like terms we get:

$$(d_1 + d_kc_k)\vec{b_1} + ... + (d_{k-1} + d_kc_{k-1})\vec{b_{k-1}} + d_kc_k\vec{b_k} + (d_{k+1} + d_kc_{k+1})\vec{b_{k+1}} + ... + (d_n + d_kc_n)\vec{b_n} = \vec{0}.$$

Now since $\vec{b_1}, ..., \vec{b_n}$ are linearly independent, all the coefficients in the above equation must be zero. In particular, we have $d_kc_k = 0$. But $c_k \neq 0$ by our hypothesis, so we may divide through by c_k and get $d_k = 0$. Now substituting 0 back in for d_k we have:

$$d_1\vec{b_1} + ...d_{k-1}\vec{b_{k-1}} + d_{k+1}\vec{b_{k+1}} + ... + d_n\vec{b_n}.$$

Now using the linear independence of $\vec{b_1}, ..., \vec{b_n}$ one more time, we conclude that $d_1 = ...d_{k-1} = d_k = d_{k+1} = ... = d_n = 0$. So the only solution is the trivial solution, and hence $\hat{\mathcal{B}}$ is linearly independent.

It remains only to check that $V = \operatorname{Span}\hat{\mathcal{B}}$. To see this, let $\vec{w} \in V$ be arbitrary. By Theorem 9, write $\vec{w}$ as a linear combination of the vectors in $\mathcal{B}$:

$$\vec{w} = a_1\vec{b_1} + ... + a_n\vec{b_n}$$

for some scalars $a_1, ..., a_n \in \mathbb{R}$. Now notice that since $\vec{v} = c_1\vec{b_1} + ...c_n\vec{b_n}$ and since $c_k \neq 0$, we may solve for the vector $\vec{b_k}$ as follows:

$$\vec{b_k} = -\tfrac{c_1}{c_k}\vec{b_1} - ... - \tfrac{c_{k-1}}{c_k}\vec{b_{k-1}} + \tfrac{1}{c_k}\vec{v} - \tfrac{c_{k+1}}{c_k}\vec{b_{k+1}} - ... - \tfrac{c_n}{c_k}\vec{b_n}.$$

Note that for the above to make sense, it is crucial that $c_k \neq 0$. Now, if we substitute the above expression for b_k in the equation $\vec{w} = a_1\vec{b_1} + ... + a_n\vec{b_n}$ and then collect like terms, we will see $\vec{w}$ written as a linear combination of the vectors $\vec{b_1}, ..., \vec{b_{k-1}}, \vec{b_{k+1}}, ..., \vec{b_n}$ and the vector $\vec{v}$. This implies that $\vec{w} \in \operatorname{Span}\hat{\mathcal{B}}$. Since $\vec{w}$ was arbitrary in V, we have $V = \operatorname{Span}\hat{\mathcal{B}}$. So $\hat{\mathcal{B}}$ is indeed a basis for V and the theorem is proved. □

Corollary 3. *Let V be a vector space which has a finite basis. Then every basis of V is the same size.*

Proof. Let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis of minimal size. Let $\mathcal{D}$ be any other basis for V, consisting of arbitrarily many elements of V. We will show that in fact $\mathcal{D}$ has exactly n elements.

Let $\vec{d_1} \in D$ be arbitrary. Since $\vec{d_1} \in V = \operatorname{Span}\{\vec{b_1}, ..., \vec{b_n}\}$, it is possible to write $\vec{d_1}$ as a nontrivial linear combination $\vec{d_1} = c_1\vec{b_1} + ... + c_n\vec{b_n}$, i.e. $c_k \neq 0$ for at least one $k \in \{1, .., n\}$. By reordering the terms of the basis $\mathcal{B}$ if necessary, we may assume without loss of generality that $c_1 \neq 0$. Define a new set $\mathcal{B}_1 = \{\vec{d_1}, \vec{b_2}, ..., \vec{b_n}\}$; by the Steinitz Exchange Lemma, $\mathcal{B}_1$ is also a basis for V.

Now we go one step further. Since $\mathcal{B}$ was a basis of minimal size, we know $\mathcal{D}$ has at least n many elements. So let $\vec{d_2} \in D$ be distinct from $\vec{d_1}$. We know $\vec{d_2} \in V = \operatorname{Span}\mathcal{B}_1$, so there exist constants $c_1, ..., c_n$ for which $\vec{d_2} = c_1\vec{d_1} + c_2\vec{b_2} + ... + c_n\vec{b_n}$. Notice that we must have $c_k \neq 0$ for some $k \in \{2, ..., n\}$, since if c_1 were the only nonzero coefficient, then $\lceil_\in$ would be a scalar multiple of $\lceil_\infty$, which is not the case since they are linearly independent. By reordering the terms $\vec{b_2}, ..., \vec{b_n}$ if necessary, we may assume without loss of generality that $c_2 \neq 0$. Define $\mathcal{B}_2 = \{\vec{d_1}, \vec{d_2}, \vec{b_3}, ..., \vec{b_4}\}$. By the Steinitz Exchange Lemma, $\mathcal{B}_2$ is also a basis for V.

Continue this process. At each step $k \in \{1, ..., n\}$, we will have obtained a new basis $\mathcal{B}_k = \{\vec{d_1}, ..., \vec{d_k}, \vec{b_{k+1}}, ..., \vec{b_n}\}$ for V. Since $\mathcal{D}$ has at least n many elements, there is a $\vec{d_{k+1}} \in V$ which is not equal to any of $\vec{d_1}, ..., \vec{d_k}$. Write $\vec{d_{k+1}} = c_1\vec{d_1} + ... + c_k\vec{d_k} + c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n}$ for some constants $c_1, ..., c_n$, and observe that one of the constants $c_{k+1}, ..., c_n$ must be nonzero since d_{k+1} is independent from $d_1, ..., d_k$. Then reorder the terms $b_{k+1}, ..., b_n$ and use the Steinitz Exchange Lemma to replace b_{k+1} with d_{k+1} to obtain a new basis.

After finitely many steps we end up with the basis $\mathcal{B}_n = \{\vec{d_1}, ..., \vec{d_n}\} \subseteq \mathcal{D}$. Since $\mathcal{B}_n$ spans V, it is not possible for $\mathcal{D}$ to contain any more elements, since nothing in V is linearly independent from the vectors in $\mathcal{B}_n$. So in fact $\mathcal{D} = \mathcal{B}_n$ and the proof is complete. □

Definition 21. Let V be a vector space, and suppose V has a finite basis $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ consisting of n vectors. Then we say that the **dimension** of V is n, or that V is **n-dimensional**. We also write that $\dim V = n$. Notice that every finite-dimensional space V has a unique dimension by Corollary 3.

If V has no finite basis, we say that V is **infinite-dimensional**.

Example 31. Determine the dimension of the following vector spaces.

(1) $\mathbb{R}$.

(2) $\mathbb{R}^2$.

(3) $\mathbb{R}^3$.

(4) $\mathbb{R}^n$.

(5) $\mathbb{P}_n$.

(6) The trivial space $\{\vec{0}\}$.

(7) Let V be the set of all infinite sequences $(a_1, a_2, a_3, ...)$ of real numbers which eventualy end in an infinite string of 0's. (Verify that V is a vector space using coordinate-wise addition and scalar multiplication.)

(8) The space $\mathcal{P}$ of all polynomials. (Verify that $\mathcal{P}$ is a vector space using standard polynomial addition and scalar multiplication.)

(9) Let V be the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$. (Verify that V is a vector space using function addition and scalar multiplication.

(10) Any line through the origin in $\mathbb{R}^3$.

(11) Any line through the origin in $\mathbb{R}^2$.

(12) Any plane through the origin in $\mathbb{R}^3$.

Corollary 4. *Suppose V is a vector space with $\dim V = n$. Then no linearly independent set in V consists of more than n elements.*

Proof. This follows from the proof of Corollary 3, since when considering $\mathcal{D}$ we only used the fact that $\mathcal{D}$ was linearly independent, not spanning. □

Corollary 5. *Let V be finite-dimensional vector space. Any linearly independent set in V may be expanded to make a basis for V.*

Proof. If a linearly independent set $\{\vec{d_1}, ..., \vec{d_n}\}$ already spans V then it is already a basis. If it doesn't span, then we can find a linearly independent vector $\vec{d_{n+1}} \notin \text{Span}\{\vec{d_1}, ..., \vec{d_n}\}$ by Fact 6 and consider $\{\vec{d_1}, ..., \vec{d_n}, \vec{d_{n+1}}\}$. If it spans V, then we're done. Otherwise, continue adding vectors. By the previous corollary, we know the process stops in finitely many steps. □

Corollary 6. *Let V be a finite-dimensional vector space. Any spanning set in V may be shrunk to make a basis for V.*

Proof. Let S be a spanning set for V, i.e. $V = \text{Span}\, S$. Without loss of generality we may assume that $\vec{0} \notin S$, for if $\vec{0}$ is in S, then we can toss it out and the set S will still span V.

If S is empty then $V = \text{Span}\, S = \{\vec{0}\}$ and S is already a basis, and we are done.

If S is not empty, then choose $\vec{s_1} \in S$. If $V = \text{Span}\{\vec{s_1}\}$, we are done and $\{\vec{s_1}\}$ is a basis. Otherwise, observe that if every vector in S were in $\text{Span}\{\vec{s_1}\}$, then we would have $\text{Span}\, S = \text{Span}\{\vec{s_1}\}$; since S spans V and $\{\vec{s_1}\}$ does not, this is impossible. So there is some $\vec{s_2} \in S$ with $\vec{s_2} \notin \text{Span}\{\vec{s_1}\}$, i.e. $\vec{s_1}$ and $\vec{s_2}$ are linearly independent.

If $\{\vec{s_1}, \vec{s_2}\}$ spans V, we are done. Otherwise we can find a third linearly independent vector $\vec{s_3}$, and so on. The process stops in finitely many steps. □

Corollary 7. *Let V be an n-dimensional vector space. Then a set of n vectors in V is linearly independent if and only if it spans V.*

11 Coordinate Systems

Definition 22. Let V be a vector space and $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for V. Let $\vec{v} \in v$. By the Unique Representation Theorem, there are unique weights $c_1, ..., c_n \in \mathbb{R}$ for which $\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n}$. Call these constants $c_1, ..., c_n$ the **coordinates of x relative to $\mathcal{B}$**. Define the **coordinate representation of x relative to $\mathcal{B}$** to be the vector

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix} \in \mathbb{R}^n.$$

Example 32. Consider a basis $\{\vec{b_1}, \vec{b_2}\}$ for $\mathbb{R}^2$, where $\vec{b_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{b_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

(1) Suppose a vector $\vec{v} \in \mathbb{R}^2$ has $\mathcal{B}$-coordinate representation $\vec{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$. Find $\vec{v}$.

(2) Let $\vec{w} = \begin{bmatrix} -1 \\ 8 \end{bmatrix}$. Find $[\vec{w}]_{\mathcal{B}}$.

Theorem 11. *Let V be a vector space and $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for V. Define a mapping $T : V \to \mathbb{R}^n$ by $T(\vec{v}) = [\vec{v}]_{\mathcal{B}}$. Then T is an isomorphism.*

Proof. (T is one-to-one.) Let $\vec{v_1}, \vec{v_2} \in V$ be such that $[\vec{v_1}]_{\mathcal{B}} = [\vec{v_2}]_{\mathcal{B}}$. Then $\vec{v_1}$ and $\vec{v_2}$ have the same coordinates relative to $\mathcal{B}$. Now the Unique Representation Theorem impies that $\vec{v_1} = \vec{v_2}$. So T is one-to-one.

(T is onto.) For every vector $\begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix} \in \mathbb{R}^n$, there corresponds a vector $\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n} \in V$. Obviously $T(\vec{v}) = \begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix}$, so the map T is onto.

(T is linear.) Let $\vec{v_1}, \vec{v_2} \in V$ be arbitrary, and let $k \in \mathbb{R}$ be an arbitrary scalar. Use the Unique Representation Theorem to find their unique coordinates $\vec{v_1} = c_1\vec{b_1} + ... + c_n\vec{b_n}$ and $\vec{v_2} = d_1\vec{b_1} + ... + d_n\vec{b_n}$. Now just check that the linearity conditions hold:

$$\begin{aligned} T(\vec{v_1} + \vec{v_2}) &= T((c_1\vec{b_1} + ... + c_n\vec{b_n}) + (d_1\vec{b_1} + ... + d_n\vec{b_n})) \\ &= T((c_1 + d_1)\vec{b_1} + ... + (c_n + d_n)\vec{b_n}) \\ &= \begin{bmatrix} c_1 + d_1 \\ ... \\ c_n + d_n \end{bmatrix} \\ &= \begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix} + \begin{bmatrix} d_1 \\ ... \\ d_n \end{bmatrix} \\ &= T(\vec{v_1}) + T(\vec{v_2}); \end{aligned}$$

and

$$\begin{aligned} T(k\vec{v_1}) &= T(kc_1\vec{b_1} + ... + kc_n\vec{b_n}) \\ &= \begin{bmatrix} kc_1 \\ ... \\ kc_n \end{bmatrix} \\ &= k\begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix} \\ &= kT(\vec{v_1}). \end{aligned}$$

This completes the proof. ∎

Corollary 8. *Any two n-dimensional vector spaces are isomorphic to one another (and to $\mathbb{R}^n$).*

Recall that isomorphisms are exactly the maps which perfectly preserve vector space structure. In other words, suppose V is a vector space and $T : \mathbb{R}^n$ is an isomorphism. Then $\vec{v_1}, .., \vec{v_n}$ are independent in V if and only if $T(\vec{v_1}), ..., T(\vec{v_n})$ are independent in $\mathbb{R}^n$; the set $\{\vec{v_1}, ..., \vec{v_n}\}$ spans V if and only if the set $\{T(\vec{v_1}), ..., T(\vec{v_n}\}$ spans $\mathbb{R}^n$; a linear combination $c_1\vec{v_1} + ... + c_n\vec{v_n}$ is equal to $\vec{w}$ in V if and only if $c_1T(\vec{v_1}) + ... + c_nT(\vec{v_n})$ is equal to $T(\vec{w})$ in W, etc., etc.

So the previous theorem and corollary imply that all problems in an n-dimensional vector space V may be effectively translated into problems in $\mathbb{R}^n$, which we generally know how to handle via matrix computations.

Example 33. Determine whether or not the vectors $1 + 2x^2$, $4 + x + 5x^2$, and $3 + 2x$ are linearly independent in $\mathbb{P}_2$.

Example 34. Determine whether or not $1 + x + x^2 \in \text{Span}\{1 + 5x, -4x + 2x^2\}$ in $\mathbb{P}_2$.

Example 35. Define a map $T : \mathbb{R}^3 \to \mathbb{P}_2$ by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = (a + b)x^2 + c$ for every $a, b, c \in \mathbb{R}$.

(1) Is T a linear transformation?
(2) Is T one-to-one?

Example 36. Let $\vec{v_1} = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}\}$. The vectors in $\mathcal{B}$ are linearly independent, so $\mathcal{B}$ is a basis for $V = \text{Span}\{\vec{v_1}, \vec{v_2}\}$. Let $\vec{w} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$. Determine if $\vec{w}$ is in V, and if it is, find $[\vec{w}]_{\mathcal{B}}$.

Fact 10 (Change of Basis: Non-Standard into Standard)**.** *Let $\vec{v} \in \mathbb{R}^n$ and let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be any basis for $\mathbb{R}^n$. Then there exists a unique matrix $P_{\mathcal{B}}$ with the property that*

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

We call $P_{\mathcal{B}}$ the ***change-of-coordinates*** *matrix from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. Moreover, we have*

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

Example 37. Let $\vec{b_1} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$, $\vec{b_2} = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}$, and $\vec{b_3} = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$. Then $\vec{b_1}, \vec{b_2}, \vec{b_3}$ are three linearly independent vectors and hence $\mathcal{B} = \{\vec{b_1}, \vec{b_2}, \vec{b_3}\}$ forms a basis for the three-dimensional space $\mathbb{R}^3$. Suppose $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$. Find $\vec{v}$.

The previous fact gives us an easy tool for converting a vector in $\mathbb{R}^n$ from its non-standard coordinate representation $[\vec{v}]_{\mathcal{B}}$ into its standard representation $\vec{v}$. What about going the opposite direction, i.e. converting from standard coordinates to non-standard? The idea will not be hard: since we can write $\vec{v} = P_{\mathcal{B}}[\vec{v}]_{\mathcal{B}}$ using a $\mathcal{B}$-to-standard matrix $P_{\mathcal{B}}$, we would like to be able to write $[\vec{v}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\vec{v}$ and regard the inverse matrix $P_{\mathcal{B}}^{-1}$ as a change-of-basis from standard coordinates to $\mathcal{B}$-coordinates. In order to justify this idea carefully, we first need to recall a few facts about inverting matrices.

12 Matrix Inversion

Definition 23. For every $n \in \mathbb{N}$, there is a unique $n \times n$ matrix I with the property that

$$AI = IA = A$$

for every $n \times n$ matrix A. This matrix is given by

$$I = \left[\begin{array}{ccc} \vec{e_1} & ... & \vec{e_n} \end{array} \right]$$

where $\vec{e_1}, ..., \vec{e_n}$ are the standard basis vectors for $\mathbb{R}^n$.

Let A be an $n \times n$ matrix. If there exists a matrix C for which $AC = CA = I$, then we say A is **invertible** and we call C the **inverse** of A. We denote $A^{-1} = C$, so

$$AA^{-1} = A^{-1}A = I.$$

Fact 11. *Let* $A = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right]$*. If* $ad - bc \neq 0$*, then* A *is invertible and*

$$A^{-1} = \frac{1}{ad - bc} \left[\begin{array}{cc} d & -b \\ -c & a \end{array} \right].$$

If $ad - bc = 0$*, then* A *is not invertible.*

Example 38. Find the inverse of $A = \left[\begin{array}{cc} 3 & 4 \\ 5 & 6 \end{array} \right]$ and verify that it is an inverse.

Example 39. Give an example of a matrix which is not invertible.

Theorem 12. *If* A *is an invertible* $n \times n$ *matrix, then for each* $\vec{w} \in \mathbb{R}^n$*, there is a unique* $\vec{v} \in \mathbb{R}^n$ *for which* $A\vec{v} = \vec{w}$*.*

Proof. Set $\vec{v} = A^{-1}\vec{w}$. Then $A\vec{v} = AA^{-1}\vec{w} = I\vec{w} = \vec{w}$. If $\vec{u}$ were another vector with the property that $A\vec{u} = \vec{w} = A\vec{v}$, then we would have $A^{-1}A\vec{u} = A^{-1}A\vec{v}$ and hence $I\vec{u} = I\vec{v}$. So $\vec{u} = \vec{v}$ and this vector is unique. $\square$

Fact 12. *Let* A *be an* $n \times n$ *matrix. If* A *is row equivalent to* I*, then* $[A|I]$ *is row equivalent to* $[I|A^{-1}]$*. Otherwise* A *is not invertible.*

Example 40. Find the inverse of $A = \left[\begin{array}{ccc} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{array} \right]$, if it exists.

Example 41. Find A^{-1} if possible.

(1) $A = \left[\begin{array}{ccc} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{array} \right]$

(2) $A = \left[\begin{array}{ccc} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{array} \right]$

13 The Connection Between Isomorphisms and Invertible Matrices

Definition 24. Let V be a vector space. Define the **identity map** with respect to V, denoted $\mathrm{id}_V : V \to V$, to be the function which leaves V unchanged, i.e. $\mathrm{id}_V(\vec{v}) = \vec{v}$ for every $\vec{v} \in V$. It is easy to verify mentally that id_V is a one-to-one and onto linear transformation, i.e. an isomorphism.

Now let W be another vector space and let $T : V \to W$ be an isomorphism. Then there is a unique **inverse map** $T^{-1} : W \to V$, with the property that

$$T^{-1} \circ T = \mathrm{id}_V \text{ and } T \circ T^{-1} = \mathrm{id}_W,$$

or

$$T^{-1}(T(\vec{v})) = \vec{v} \text{ for every } \vec{v} \in V \text{ and } T(T^{-1}(\vec{w})) = \vec{w} \text{ for every } \vec{w} \in W.$$

Example 42. Let $T : \mathbb{R}^3 \to \mathbb{P}_2$ be defined by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = ax^2 + bx + c$. We proved T is an isomorphism on the homework. Find an expression for T^{-1}.

Theorem 13. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an isomorphism and let A be its unique $n \times n$ matrix representation as guaranteed by 3. Then A is invertible and the matrix representation of T^{-1} is A^{-1}.*

Proof. First notice that if T is an isomorphism, it is one-to-one and hence the columns of A are linearly independent by Fact 5. This means that A may be row reduced to the $n \times n$ identity matrix I, and hence A is invertible by Fact 12.

Notice that the unique $n \times n$ matrix representation of the identity map $\mathrm{id}_{\mathbb{R}^n}$ must be the $n \times n$ identity matrix I. It is sufficient to notice that $\mathrm{id}_{\mathbb{R}^n}(\vec{e_k}) = \vec{e_k} = I\vec{e_k}$ for every standard basis vector $\vec{e_k}$, $1 \leq k \leq n$.

Now let $\vec{v} \in \mathbb{R}^n$ be arbitrary. Set $\vec{u} = A^{-1}\vec{v}$. Notice that

$$T(\vec{u}) = A\vec{u} = A(A^{-1}\vec{v}) = (AA^{-1})\vec{v} = I\vec{v} = \vec{v}.$$

Taking T^{-1} of both sides above, we get $T^{-1}(\vec{v}) = \vec{u} = A^{-1}\vec{v}$. So T^{-1} and A^1 send $\vec{v}$ to the same place for every $\vec{v}$, i.e. A^{-1} is a matrix representation for T^{-1}. Since this matrix representation must be unique, we are done. $\square$

Example 43. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation defined by the rules $T(\vec{e_1}) = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ and $T(\vec{e_2}) = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$.

(1) Prove that T is an isomorphism.
(2) Find an explicit formula for T^{-1}.

Solution. (1) T has matrix representation $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$ by Theorem 3. A is invertible by Fact 11 since $3 \cdot 6 - 4 \cdot 5 \neq 0$. So T is invertible, i.e. T is an isomorphism.

(2) By Fact 11, we have $A^{-1} = -\frac{1}{2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix}$. Then by Theorem 13, we have $T(\vec{v}) = -\frac{1}{2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix}\vec{v}$ for every $\vec{v} \in \mathbb{R}^2$. $\square$

14 More Change of Basis

Fact 13 (Change of Basis: Standard into Non-Standard)**.** *Let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a basis for $\mathbb{R}^n$. Let $P_{\mathcal{B}}$ be the unique $n \times n$ for which*

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

for every $\vec{v} \in \mathbb{R}^n$, so $P_{\mathcal{B}}$ is the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis. Then $P_{\mathcal{B}}$ is always invertible since its columns are linearly independent, and we call the inverse $P_{\mathcal{B}}^{-1}$ the change-of-coordinates matrix from the standard basis to $\mathcal{B}$. This matrix has the property that

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

for every $\vec{v} \in \mathbb{R}^n$.

Example 44. Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$. Find a matrix A such that $[\vec{v}]_{\mathcal{B}} = A\vec{v}$ for every $\vec{v} \in \mathbb{R}^2$.

Example 45. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} -7 \\ 9 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \end{bmatrix} \right\}$. Find a matrix A for which $[\vec{v}]_{\mathcal{C}} = A[\vec{v}]_{\mathcal{B}}$ for every $\vec{v} \in \mathbb{R}^2$.

Fact 14. *Let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ and $\mathcal{C} = \{\vec{c_1}, ..., \vec{c_n}\}$ be two bases for $\mathbb{R}^n$. Then there exists a unique $n \times n$ matrix, which we denote $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, with the property that*

$$[\vec{v}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\vec{v}]_{\mathcal{B}}$$

for every $\vec{v} \in \mathbb{R}^n$. We call $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.

Notice that

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = P_{\mathcal{C}}^{-1} \cdot P_{\mathcal{B}}.$$

We may also write

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} \vec{a_1} & ... & \vec{a_n} \end{bmatrix}.$$

where $\vec{a_1} = [\vec{b_1}]_{\mathcal{C}}$, ..., $\vec{a_n} = [\vec{b_n}]_{\mathcal{C}}$.

Fact 15. *Let $\mathcal{B}$ and $\mathcal{C}$ be any two bases for $\mathbb{R}^n$. Then $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is invertible, and $\left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$.*

Example 46. Let $\mathcal{B}$ and $\mathcal{C}$ be as in the previous example.

(1) Find $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$.
(2) Find $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$.

Example 47. Suppose $\mathcal{C} = \{\vec{c_1}, \vec{c_2}\}$ is a basis for $\mathbb{R}^2$, and $\mathcal{B} = \{\vec{b_1}, \vec{b_2}\}$ where $\vec{b_1} = 4\vec{c_1} + \vec{c_2}$ and $\vec{b_2} = -6\vec{c_1} + \vec{c_2}$. Suppose $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find $[\vec{v}]_{\mathcal{C}}$.

15 Rank and Nullity

Definition 25. Let V and W be vector spaces and $T : V \to W$ a linear transformation. Define the **rank** of T, denoted $\operatorname{rank} T$, by

$$\operatorname{rank} T = \dim \operatorname{Ran} T.$$

We also define the **nullity** of T to be $\dim \operatorname{Nul} T$, the dimension of the null space of T.

Intuitively we think of the nullity of a linear transformation T as being the amount of "information lost" by the map, and the rank is the amount of "information preserved," as the next example may help illustrate.

Example 48. Compute $\operatorname{rank} T$ and $\dim \operatorname{Nul} T$ for the following maps.

(1) $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} y \\ z \end{bmatrix}$.

(2) $T : \mathbb{R}^3 \to \mathbb{R}^3$, $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ 2x+2y \\ 0 \end{bmatrix}$.

(3) $T : \mathbb{R} \to \mathbb{R}^2$, $T(x) = \begin{bmatrix} -2x \\ x \end{bmatrix}$.

(4) $T : \mathbb{R}^5 \to \mathbb{R}^3$, $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Theorem 14 (Rank-Nullity Theorem)**.** *Let V, W be finite-dimensional vector spaces and let $T : V \to W$ be a linear transformation. Then*

$$\operatorname{rank} T + \dim \operatorname{Nul} T = \dim V.$$

Proof. Suppose $\dim \operatorname{Nul} T = k$ and $\dim V = n$. Obviously $k \leq n$ since $\operatorname{Nul} T$ is a subspace of V. Let $\{\vec{b_1}, ..., \vec{b_k}\}\}$ be a basis for $\operatorname{Nul} T$. By Corollary 5 we may expand this set to form a basis $\{\vec{b_1}, ..., \vec{b_k}, \vec{b_{k+1}}, ..., \vec{b_n}\}$ for the larger space V. We claim that the set $\{T(\vec{b_{k+1}}), ..., T(\vec{b_n})\}$ comprises a basis for $\operatorname{Ran} T$, and hence $\operatorname{rank} T = \dim \operatorname{Ran} T = n - k$.

First we must show that $T(\vec{b_{k+1}}), ..., T(\vec{b_n})$ are linearly independent. To see this, consider the equation

$$c_{k+1}T(\vec{b_{k+1}}) + ... + c_nT(\vec{b_n}) = \vec{0}.$$

Since T is a linear transformation, we may rewrite the above as

$$T(c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n}) = \vec{0}.$$

It follows that the linear combination $c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n}$ is in the null space $\operatorname{Nul} T$. This means we can write it as a linear combination of the elements of our basis $\{\vec{b_1}, ..., \vec{b_k}\}$ for $\operatorname{Nul} T$:

$$c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n} = c_1\vec{b_1} + ... + c_k\vec{b_k}$$

for some $c_1, ..., c_k \in \mathbb{R}$. Moving all terms to one side of the equation, we get:

$$c_1\vec{b_1} + ... + c_k\vec{b_k} - c_{k+1}\vec{b_{k+1}} - ... - c_n\vec{b_n} = \vec{0}.$$

In other words we have written $\vec{0}$ as a linear combination of the basis elements $\{\vec{b_1}, ..., \vec{b_n}\}$ for V. Since these are linearly independent, all the coefficients must be 0, i.e. $c_1 = ... = c_k = c_{k+1} = ... = c_n = 0$. This shows that $T(\vec{b_{k+1}}), ..., T(\vec{b_n})$ are linearly independent.

It remains only to show that $\{T(\vec{b_{k+1}}), ..., T(\vec{b_n})\}$ forms a spanning set for $\operatorname{Ran} T$. To that end, let $\vec{w} \in \operatorname{Ran} T$, so there exists a $\vec{v} \in V$ with $T(\vec{v}) = \vec{w}$. Write $\vec{v}$ as a linear combination of the basis elements for V:

$$\vec{v} = c_1\vec{b_1} + ... + c_k\vec{b_k} + c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n}.$$

Notice that $T(\vec{b_1}) = ... = T(\vec{b_k}) = \vec{0}$ since $\vec{b_1}, ..., \vec{b_k} \in \operatorname{Nul} T$. So, applying the map T to both sides of the equation above, we get:

$$\begin{aligned}
\vec{w} &= T(\vec{v}) \\
&= T(c_1\vec{b_1} + ... + c_k\vec{b_k} + c_{k+1}\vec{b_{k+1}} + ... + c_n\vec{b_n}) \\
&= c_1T(\vec{b_1}) + ... + c_kT(\vec{b_k}) + c_{k+1}T(\vec{b_{k+1}}) + ... + c_nT(\vec{b_n}) \\
&= c_1\vec{0} + ... + c_k\vec{0} + c_{k+1}T(\vec{b_{k+1}}) + ... + c_nT(\vec{b_n}) \\
&= c_{k+1}T(\vec{b_{k+1}}) + ... + c_nT(\vec{b_n}).
\end{aligned}$$

So we have written $\vec{w}$ as a linear combination of the elements of $\{T(\vec{b_{k+1}}), ..., T(\vec{b_n})\}$, and since $\vec{w}$ was arbitrary, we see that they span $\operatorname{Ran} T$. So $\{T(\vec{b_{k+1}}), ..., T(\vec{b_n})\}$ is a basis for $\operatorname{Ran} T$ and hence $\operatorname{rank} T = n - k$. The conclusion of the theorem now follows. □

Theorem 15. *Let V and W both be n-dimensional vector spaces, and let $T : V \to W$ be a linear transformation. Then the following statements are all equivalent.*

(1) T is an isomorphism.
(2) T is one-to-one and onto.
(3) T is invertible.
(4) T is one-to-one.
(5) T is onto.
(6) $\operatorname{Nul} T = \{\vec{0}\}$.
(7) $\dim \operatorname{Nul} T = 0$.
(8) $\operatorname{Ran} T = W$.
(9) $\operatorname{rank} T = n$.

Furthermore, if $V = W = \mathbb{R}^n$ and A is the unique $n \times n$ matrix representation of T, then the following statements are all equivalent, and equivalent to the statements above.

(10) A is invertible.
(11) A is row-equivalent to the $n \times n$ identity matrix I.
(12) The equation $A\vec{x} = \vec{0}$ has only the trivial solution.
(13) The columns of A form a basis for $\mathbb{R}^n$.
(14) The columns of A are linearly independent.
(15) The columns of A span $\mathbb{R}^n$.
(16) $\operatorname{Col} A = \mathbb{R}^n$.
(17) $\dim \operatorname{Col} A = n$.

Proof. We first check that (5) $\Leftrightarrow$ (8) $\Leftrightarrow$ (9). We have (5) $\Leftrightarrow$ (8) by Theorem 7. If (8) holds and $\operatorname{Ran} T = W$, then $\operatorname{rank} T = \dim \operatorname{Ran} T = \dim W = n$, so (9) holds. Conversely if (9) holds, then $\operatorname{Ran} T$ admits a basis $\mathcal{B}$ consisting of n vectors. The vectors in $\mathcal{B}$ are linearly independent in W and hence they span W by Corollary 7. So $W = \operatorname{Span} \mathcal{B} = \operatorname{Ran} T$ and hence (8) holds; so (5), (8), and (9) are equivalent.

We also check that (4) $\Leftrightarrow$ (6) $\Leftrightarrow$ (7). (4) $\Leftrightarrow$ (6) follows from Theorem 5 and (6) $\Leftrightarrow$ (7) is obvious from the definition of dimension.

By the Rank-Nullity Theorem 14, we have (7) $\Leftrightarrow$ (9). So all of statements (4) − (9) are equivalent. Statements (1) − (3) are obviously equivalent, and obviously imply (4) − (9). If any of statements (4) − (9) hold, then all of them do- in particular, T will be one-to-one and onto, so (1) − (3) will hold. This concludes the proof of the first half of the theorem.

Now suppose we have $V = W = \mathbb{R}^n$ and A is the $n \times n$ matrix representation of T. We have (1) $\Rightarrow$ (10) by Theorem 13. If (10) holds and A is invertible, then we may define a map $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ by $T^{-1}(\vec{v}) = A^{-1}\vec{v}$ for every $\vec{v} \in \mathbb{R}^n$; it is easy to check that this T^{-1} is truly the inverse of T, and hence (3) holds. So (10) holds if and only if (1) − (9) holds.

We have (10) $\Leftrightarrow$ (11) by Fact 12. (11) holds if and only if A is row-equivalent to I, if and only if the augmented matrix $[A \mid \vec{0}]$ is row-equivalent to $[I \mid \vec{0}]$, if and only if the equation $A\vec{x} = \vec{0}$ has only one solution, i.e. (12) holds. So (11) $\Leftrightarrow$ (12). So (1) − (12) are all equivalent.

Since $\operatorname{Ran} T = \operatorname{Col} A$ by Theorem 8, we have (16) $\Leftrightarrow$ (8) and (17) $\Leftrightarrow$ (9). So (1) − (12) are each equivalent to (16) and (17).

We have (15) $\Leftrightarrow$ (16) by the definition of $\operatorname{Col} A$, and we have (13) $\Leftrightarrow$ (14) $\Leftrightarrow$ (15) by Corollary 7. So (1) − (17) are all equivalent as promised. □

16 Determinants and the LaPlace Cofactor Expansion

Theorem 15 says that linear transformations are actually isomorphisms whenever their matrix representations are invertible. When is a given $n \times n$ matrix invertible? We have easy ways to check if $n = 1$ or $n = 2$, but how can one tell if, say, a 52×52 matrix is invertible? We hope to develop a computable formula for checking the invertibility of an arbitrary square matrix, which we will call the determinant of that matrix.

There are many different ways to introduce and define the notion of a determinant for a general $n \times n$ matrix A. In the interest of brevity, we will restrict our attention here to two major approaches. The first method of computing a determinant is given implicitly in the upcoming Definition 28. This definition best conveys the intuition that the determinant of a matrix characterizes its invertibility or non-invertibility, and is computationally easiest for large matrices. The second method for computing determinants is the LaPlace expansion or cofactor expansion, which is probably computationally easier for small matrices, and which will prove very useful in the next section on eigenvalues and eigenvectors.

Definition 26. (1) Let $A = [a]$ be a 1×1 matrix. The **determinant** of A, denoted $\det A$, is the number $\det A = a$.

(2) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix. The **determinant** of A, also denoted $\det A$, is the number $\det A = ad - bc$.

Notice that in both cases above, the matrix A is invertible if and only if $\det A \neq 0$. This statement is obvious for (1), and for (2) it follows from Fact 11, which the reader may be asked to prove on the homework.

We wish to generalize this idea for $n \times n$ matrices A of arbitrary size n, i.e. we hope to find some computable number $\det A$ with the property that A is invertible if and only if $\det A \neq 0$. Of course A is invertible if and only if it is row-equivalent to the $n \times n$ identity matrix I, so there must be a strong relationship between the determinant $\det A$ and the elementary row operations we use to reduce a matrix. Building on this vague observation, in the next example we investigate the effect of the row operations on $\det A$ for 2×2 matrices A.

Example 49. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix.

(1) If $c = 0$, then what is $\det A$?

(2) Row-reduce A to echelon form without interchanging any rows and without scaling any rows. Let E be the new matrix obtained after this reduction. What is $\det E$?

(3) Let $k \in \mathbb{R}$. What is $\det \begin{bmatrix} ka & kb \\ c & d \end{bmatrix}$? What about $\det \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$? What about $\det \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}$?

(4) What is $\det \begin{bmatrix} c & d \\ a & b \end{bmatrix}$?

Definition 27. Let A be an $n \times n$ matrix. Write

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \dots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

so a_{ij} denotes the entry of A in row i and column j. For shorthand we also denote this matrix by $A = [a_{ij}]$.

A matrix $A = [a_{ij}]$ is called **upper-triangular** if $a_{ij} = 0$ whenever $j < i$. A matrix is obviously upper-triangular if and only if it is in echelon form.

Definition 28. We postulate the existence of a number $\det A$ associated to any matrix $A = [a_{ij}]$, called the **determinant** of A, which satisfies the following properties.

(1) If A is upper-triangular then $\det A = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}$, i.e. the determinant is the product of the entries along the diagonal of A.

(2) (Replacement) If E is a matrix obtained by using a row replacement operation on A, then $\det A = \det E$.

(3) (Interchange) If E is a matrix obtained by using the row interchange operation on A, then $\det A = -\det E$.

(4) (Scaling) If E is a matrix obtained by scaling a row of A by some constant $k \in \mathbb{R}$, then $\det A = \frac{1}{k} \det E$.

Notice that the definition above is consistent with our definitions of $\det A$ for 1×1 and 2×2 matrices A. The definition above gives a recursive, or algorithmic method for computing determinants of larger matrices A, as we will see in the upcoming examples. The fact that the above definition is a good one, i.e. that such a number $\det A$ exists for each $n \times n$ matrix A and is also unique, for every $n > 2$, should not be at all obvious to the reader, but its proof would involve a long digression and so unfortunately we must omit it here.

Example 50. (1) Compute $\det \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 & 9 \\ 0 & 0 & -1 & -2 & -3 \\ 0 & 0 & 0 & -4 & -5 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$.

(2) Compute $\det \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$.

(3) Compute $\det \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

Theorem 16. *A square matrix is invertible if and only if* $\det A \neq 0$.

Proof. ($\Rightarrow$) Suppose $\det A \neq 0$. Notice that by Definition 28, if E is the matrix obtained after performing any elementary row operation on A, then we either have $\det E = \det A$, or $\det E = -\det A$, or $\det E = k \det A$ for some constant $k \in \mathbb{R}$. In any of the three cases $\det E \neq 0$ since $\det A \neq 0$.

It follows that if $B = [b_{ij}]$ is the echelon form of A, then $\det B \neq 0$, since B is obtained from A by a finite sequence of elementary row operations. But $\det B = b_{11} \cdot b_{22} \cdot ... \cdot b_{nn}$ since B is upper-triangular. Since $\det B \neq 0$, we have $b_{11} \neq 0$, $b_{22} \neq 0$, ..., $b_{nn} \neq 0$. So all the entries along the diagonal are non-zero, and hence A is row-equivalent to B is row-equivalent to the identity matrix I. Hence A is invertible by Theorem 15.

($\Leftarrow$) On the other hand suppose $\det A = 0$, and let E be the matrix obtained after using any elementary row operation on A. Then either $\det E = \det A$ or $\det E = -\det A$ or $\det E = k \det A$ for some constant k; in all three cases $\det E = 0$ since $\det A = 0$. Then if B is the echelon form of A, we have $\det B = 0$ since B is obtained from A after applying finitely many elementary row operations. It follows that $\det B = b_{11} \cdot b_{22} \cdot ... \cdot b_{nn} = 0$ and hence $b_{ii} = 0$ for some $i \in \{1, ..., n\}$. So B is not row-equivalent to I, and hence neither is A. Then A is not invertible by Theorem 15. $\square$

Definition 29. Let $A = [a_{ij}]$ be an $n \times n$ matrix. For any fixed integers i and j, let A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column from A. Written explicitly, we have:

$$A_{ij} = \begin{bmatrix} a_{11} & ... & a_{1,i-1} & a_{1,i+1} & ... & a_{1n} \\ & & ... & & & \\ a_{j-1,1} & ... & a_{j-1,i-1} & a_{j-1,i+1} & ... & a_{j-1,n} \\ a_{j+1,1} & ... & a_{j+1,i-1} & a_{j+1,i+1} & ... & a_{j+1,n} \\ & & ... & & & \\ a_{n1} & ... & a_{n,j-1} & a_{n,j+1} & ... & a_{nn} \end{bmatrix}.$$

The (i, j)**-cofactor** of A is the number

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

Fact 16 (LaPlace Expansion for Determinants)**.** *Let* $A = [a_{ij}]$ *be an* $n \times n$ *matrix. Then for any* $i, j \in \{1, ..., n\}$,

$$\begin{aligned} \det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + ... + a_{in}C_{in} \\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + ... + a_{nj}C_{nj}. \end{aligned}$$

The above is also called the method of **cofactor expansion** for computing determinants.

Example 51. Use a cofactor expansion across the third row to compute $\det A$, where $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$.

Example 52. Compute $\det A$, where

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

17 Eigenvectors and Eigenvalues

Definition 30. Let V be a vector space and $T : V \to V$ a linear transformation. A number λ is called a **eigenvalue** of T if there exists a non-zero vector $\vec{v} \in V$ for which

$$T(\vec{v}) = \lambda\vec{v}.$$

If λ is an eigenvalue of T, then any vector $\vec{v}$ for which $T(\vec{v}) = \lambda\vec{v}$ is called a **eigenvector** of T corresponding to the eigenvalue λ.

Example 53. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which has matrix representation $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

(1) Plot the vectors $T(\vec{e_1})$, $T(\vec{e_2})$, and $T(\vec{e_1} + \vec{e_2})$ in the Cartesian coordinate plane.
(2) Which of $\vec{e_1}$, $\vec{e_2}$, and $\vec{e_1} + \vec{e_2}$ are eigenvectors of T? For what eigenvalues λ?

Example 54. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which has matrix representation $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$.

(1) Show that 7 is an eigenvalue of T.
(2) Find all eigenvectors which correspond to the eigenvalue 7.

Definition 31. Let V be a vector space and $T : V \to V$ a linear transformation. Let λ be an eigenvalue of T. The set of all eigenvectors corresponding to λ is called the **eigenspace** of T corresponding to λ.

Theorem 17. *Let V be a vector spaces, and $T : V \to V$ a linear transformation. Let λ be an eigenvalue of T. Let $\mathrm{id}_V : V \to V$ be the identity map as in Definition 24. Then the eigenspace of T corresponding to λ is exactly $\mathrm{Nul}(T - \lambda \cdot \mathrm{id}_V)$, where $T - \lambda \cdot \mathrm{id}_V : V \to V$ is defined by $(T - \lambda \cdot \mathrm{id}_V)(\vec{v}) = T(\vec{v}) - \lambda \cdot \mathrm{id}_V(\vec{v}) = T(\vec{v}) - \vec{v}$ for every $\vec{v} \in V$.*

Proof. Suppose $\vec{v} \in V$ is an eigenvalue of T corresponding to λ, i.e.

$$T(\vec{v}) = \lambda\vec{v}.$$

Then moving all terms to the left side of the equality, we have

$$T(\vec{v}) - \lambda\vec{v} = \vec{0}.$$

Rewriting the above, we have

$$(T - \lambda \cdot \mathrm{id}_V)(\vec{v}) = T(\vec{v}) - \lambda \cdot \mathrm{id}_V(\vec{v}) = T(\vec{v}) - \lambda\vec{v} = \vec{0}.$$

So $\vec{v} \in \mathrm{Nul}(T - \lambda \cdot \mathrm{id}_V)$. The same arguments above, applied in reverse, will show that if $\vec{v} \in \mathrm{Nul}(T - \lambda \cdot \mathrm{id}_V)$, then $T(\vec{v}) = \lambda\vec{v}$ and hence $\vec{v}$ is an eigenvector corresponding to λ. So the eigenspace corresponding to λ is exactly $\mathrm{Nul}(T - \lambda \cdot \mathrm{id}_V)$. □

Corollary 9. *Let V be a vector space and $T : V \to V$ a linear transformation. If λ is any eigenvalue of T, then its corresponding eigenspace is a vector subspace of V.*

Corollary 10. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be its $n \times n$ matrix representation. Let λ be an eigenvalue of T, and let I denote the $n \times n$ identity matrix. Then the eigenspace corresponding to λ is exactly $\mathrm{Nul}(A - \lambda I)$.*

Example 55. Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

Example 56. Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

Example 57. Find the eigenvalues of $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & -2 \\ -1 & 2 & 3 \end{bmatrix}$.

Definition 32. Let A be an $n \times n$ matrix. Then the **characteristic polynomial** of A is the function

$$\det(A - \lambda I),$$

where λ is treated as a variable.

Theorem 18. *Let A be an $n \times n$ matrix. A number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.*

Proof. λ is an eigenvalue if and only if there is a non-zero vector $\vec{v}$ with $A\vec{v} = \lambda\vec{v}$. This happens if and only if there is a non-zero vector $\vec{v}$ for which $(A - \lambda I)\vec{v} = \vec{0}$, i.e. if and only if $\mathrm{Nul}(A - \lambda I)$ is non-trivial. $\mathrm{Nul}(A - \lambda I)$ is non-trivial if and only if $A - \lambda I$ is not invertible, by Theorem 15. But $A - \lambda I$ is not invertible if and only if $\det(A - \lambda I) = 0$ by Theorem 16. □

Example 58. Let $A = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 4 & -1 \\ 1 & 0 & 2 \end{bmatrix}$.

(1) Find the characteristic polynomial of A.

(2) Find all eigenvalues of A. (*Hint:* 3 is one.)

(3) For each eigenvalue λ of A, find a basis for the eigenspace corresponding to λ.

18 Diagonalization

Definition 33. An $n \times n$ matrix $A = [a_{ij}]$ is called **diagonal** if $a_{ij} = 0$ whenever $i \neq j$, i.e. the only non-zero entries in A are along the main diagonal.

We think of diagonal matrices as the "simplest" of all possible linear transformations. To illustrate this intuition, consider the following example.

Example 59. Let $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$.

(1) Find a formula for the matrix power A^k.

(2) Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Observe that $A = PDP^{-1}$, where $P = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$. Find a formula for the matrix power A^k.

Definition 34. Let A and C be $n \times n$ matrices. We say that A is **similar** to C if there exists an invertible $n \times n$ matrix P such that $A = PCP^{-1}$.

Here is some intuition for the above definition. Let $\mathcal{E}$ be the standard basis for $\mathbb{R}^n$. Since P is invertible, its columns form a basis $\mathcal{B}$ for $\mathbb{R}^n$ and hence $P = P_{\mathcal{B}} = \underset{\mathcal{E} \leftarrow \mathcal{B}}{P}$ is the change-of-basis matrix from $\mathcal{B}$ to the standard basis. Likewise $P^{-1} = P_{\mathcal{B}}^{-1} = \underset{\mathcal{B} \leftarrow \mathcal{E}}{P}$ is the change-of-basis matrix from the standard basis to $\mathcal{B}$. Then the following is a diagram of the transformations A, C, P, and P^{-1}:

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^n \\ P^{-1}\downarrow & & P\uparrow \\ \mathbb{R}^n & \xrightarrow{C} & \mathbb{R}^n \end{array}$$

The two "up-down" arrows above represent just a change of basis. To say that C is similar to B is to say that travelling along the top arrow is the same as travelling along the bottom three arrows. In other words, the matrix C acting on basis $\mathcal{B}$-coordinates induces the same transformation of $\mathbb{R}^n$ as A does acting on standard coordinates.

Definition 35. An $n \times n$ matrix A is called **diagonalizable** if A is similar to a diagonal matrix D.

Theorem 19 (Diagonalization Theorem)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.*

Moreover, if $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ is a set of n linearly independent eigenvectors of A corresponding to the eigenvalues $\lambda_1, ..., \lambda_n$ respectively, then

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

where $P = P_{\mathcal{B}}$ and D is a diagonal matrix with entries $\lambda_1, ..., \lambda_n$ along the diagonal.

Proof. ($\Rightarrow$) Let $\mathcal{B} = \{\vec{b_1}, ..., \vec{b_n}\}$ be a collection of n linearly independent eigenvectors of A corresponding to the eigenvalues $\lambda_1, ..., \lambda_n$ respectively. Set $P = P_{\mathcal{B}}$ and let D be the diagonal matrix with entries λ_1,...,λ_n along the diagonal. We can use two different arguments to show that $A = PDP^{-1}$; the first is purely computational while the second uses the fact that $\mathcal{B}$ is a basis for $\mathbb{R}^n$.

(Argument 1) Recall that $P_{\mathcal{B}} = \begin{bmatrix} \\ \vec{b_1} & ... & \vec{b_n} \\ \\ \end{bmatrix}$. In that case, we have

$$AP = A\begin{bmatrix} \\ \vec{b_1} & ... & \vec{b_n} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ A\vec{b_1} & ... & A\vec{b_n} \\ \\ \end{bmatrix}$$

and

$$PD = \begin{bmatrix} \\ \vec{b_1} & ... & \vec{b_n} \\ \\ \end{bmatrix}\begin{bmatrix} \lambda_1 & ... & 0 \\ & ... & \\ 0 & ... & \lambda_n \end{bmatrix} = \begin{bmatrix} \\ \lambda_1\vec{b_1} & ... & \lambda_n\vec{b_n} \\ \\ \end{bmatrix}.$$

Since $A\vec{b_1} = \lambda_1\vec{b_1}$, ..., $A\vec{b_n} = \lambda_n\vec{b_n}$, we have that $AP = PD$. Now multiplying on the right by P^{-1}, we get $A = PDP^{-1}$.

(Argument 2) We will show that $A\vec{v} = PDP^{-1}\vec{v}$ for every vector $\vec{v} \in \mathbb{R}^n$, and hence $A = PDP^{-1}$ by the uniqueness clause in Theorem 3. So let $\vec{v} \in \mathbb{R}^n$ be arbitrary. Since $\mathcal{B}$ is a basis for $\mathbb{R}^n$, there exist constants $c_1, ..., c_n$ for which $\vec{v} = c_1\vec{b_1} + ... + c_n\vec{b_n}$. Now compute $A\vec{v}$:

$$\begin{aligned} A\vec{v} &= A(c_1\vec{b_1} + ... + c_n\vec{b_n}) \\ &= c_1A\vec{b_1} + ... + c_nA\vec{b_n} \\ &= c_1\lambda_1\vec{b_1} + ... + c_n\lambda_n\vec{b_n}. \end{aligned}$$

To compute $PDP^{-1}\vec{v}$, observe that P^{-1} is the change-of-basis matrix from the standard basis to $\mathcal{B}$; so $P^{-1}\vec{v} = \begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix}$. Since D is a diagonal matrix with entries $\lambda_1, ..., \lambda_n$, we have $AP^{-1}\vec{v} = A\begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix} = \begin{bmatrix} \lambda_1 c_1 \\ ... \\ \lambda_n c_n \end{bmatrix}$. Lastly since P is the change-of-basis matrix from $\mathcal{B}$ to the standard basis, we have $PDP^{-1}\vec{v} = P\begin{bmatrix} \lambda_1 c_1 \\ ... \\ \lambda_n c_n \end{bmatrix} = \lambda_1 c_1 \vec{b_1} + ... + \lambda_n c_n \vec{b_n} = A\vec{v}$. Since $\vec{v}$ was arbitrary, this completes the proof.

($\Leftarrow$) Conversely, suppose A is diagonalizable, i.e. $A = PDP^{-1}$ for some diagonal matrix D and some invertible matrix P. Write $P = \begin{bmatrix} \vec{b_1} & ... & \vec{b_n} \end{bmatrix}$, and denote the entries along the diagonal of D by $\lambda_1, ..., \lambda_n$ respectively. By multiplying on the right by P, we see that $AP = PD$ and hence

$$\begin{bmatrix} A\vec{b_1} & ... & A\vec{b_n} \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{b_1} & ... & \lambda_n\vec{b_n} \end{bmatrix}$$

as we computed previously. So all columns are equal, i.e. $A\vec{b_1} = \lambda_1\vec{b_1}$, ..., $A\vec{b_n} = \lambda_n\vec{b_n}$. This implies that $\vec{b_1}, ..., \vec{b_n}$ are eigenvectors of A corresponding to $\lambda_1, ..., \lambda_n$ respectively. In addition since P is invertible, its columns $\vec{b_1}, ..., \vec{b_n}$ are linearly independent. So A has n linearly independent eigenvectors. $\square$

Example 60. Diagonalize the following matrix, if possible.

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

Example 61. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.

Definition 36. Recall from a previous course that if $p(x)$ is a polynomial and a is a root of p (i.e. $p(a) = 0$), then the **multiplicity** of a with respect to p is the number of times the term $(x - a)$ appears in the factored form of p. (This factored form always exists by the Fundamental Theorem of Algebra.) For an example, if $p(x) = 5(x-2)^4(x-3)^5(x-4)$, then p has three roots 2, 3, and 4, and the roots have multiplicity 4, 5, and 1 respectively. The sum of the multiplicities is $4 + 5 + 1 = 10$, which is exactly the degree of p.

Now let A be an $n \times n$ matrix and let λ_1 be an eigenvalue of A. Then define the **multiplicity** of λ_1 with respect to A to be just the multiplicity of λ_1 with respect to the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$.

Fact 17. *Let A be an $n \times n$ matrix. Let $\lambda_1, ..., \lambda_p$ be all the distinct eigenvalues of A, with multiplicities $m_1, ..., m_p$ respectively. Then $m_1 + ... + m_p \leq n$.*

The reason we have a $\leq$ above instead of an $=$ is because for the moment, we are considering only real eigenvalues and not complex eigenvalues. For example, the characteristic polynomial of $A = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$ is $p(\lambda) = \lambda^2 + 2\lambda + 2$, which is an irreducible quadratic, i.e. it cannot be factored into a product of degree one binomials with real coefficients. So A has no real eigenvalues and hence the sum of the multiplicities of its eigenvalues is 0, which is strictly less than 2. (If we allowed complex eigenvalues, we would always have $m_1 + ... + m_p = n$ in the above fact.)

Theorem 20. *Let A be an $n \times n$ matrix. Let $\lambda_1, ..., \lambda_p$ be all the distinct eigenvalues of A, with multiplicities $m_1, ..., m_p$ respectively.*

(1) *For $1 \leq k \leq p$, the dimension of the eigenspace corresponding to λ_k is less than or equal to m_k.*

(2) *A is diagonalizable if and only if the sum of the dimensions of the eigenspaces is equal to n, and this happens if and only if the dimension of the eigenspace corresponding to λ_k is m_k for all $1 \leq k \leq p$.*

(3) *If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each $1 \leq k \leq p$, then the total collection of vectors in $\mathcal{B}_1, ..., \mathcal{B}_p$ comprises a basis of eigenvectors for $\mathbb{R}^n$.*

(4) *If $p = n$ then A is diagonalizable.*

Example 62. Diagonalize the following matrix, if possible.

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